

# **Inquiry: The University of Arkansas Undergraduate Research Journal**

---

Volume 7

Article 1

---

Fall 2006

## **Inquiry: The University of Arkansas Undergraduate Research Journal - Volume 7 - Fall 2006**

Inquiry Editors

Follow this and additional works at: <http://scholarworks.uark.edu/inquiry>

---

### **Recommended Citation**

Editors, Inquiry (2006) "Inquiry: The University of Arkansas Undergraduate Research Journal - Volume 7 - Fall 2006," *Inquiry: The University of Arkansas Undergraduate Research Journal*: Vol. 7 , Article 1.

Available at: <http://scholarworks.uark.edu/inquiry/vol7/iss1/1>

This Entire Issue is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Inquiry: The University of Arkansas Undergraduate Research Journal by an authorized editor of ScholarWorks@UARK. For more information, please contact [scholar@uark.edu](mailto:scholar@uark.edu).

THE UNIVERSITY OF ARKANSAS

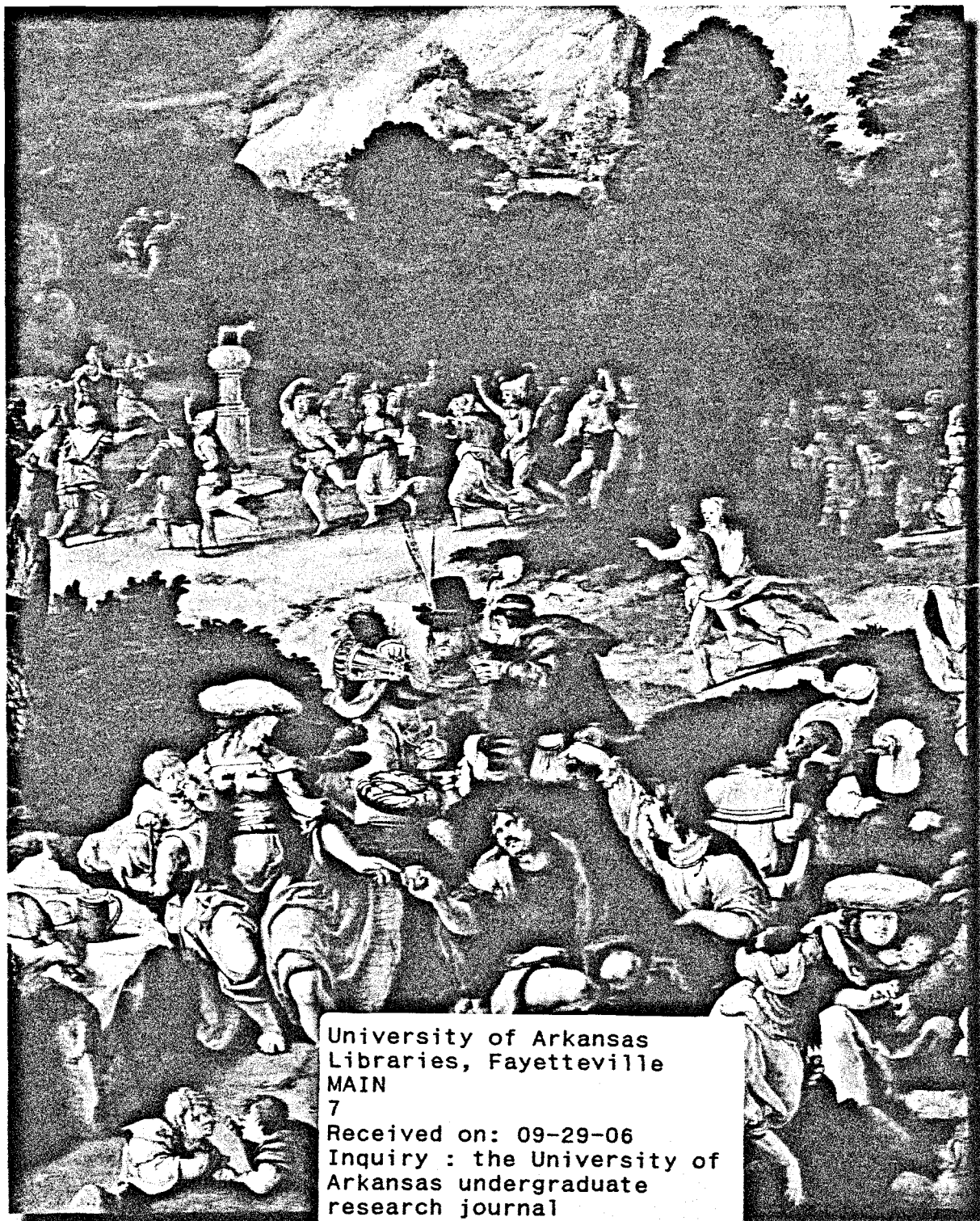
UNDERGRADUATE RESEARCH JOURNAL

Editors: Inquiry: The University of Arkansas Undergraduate Research Journal

# INQUIRY

Volume 7

Fall 2006



University of Arkansas  
Libraries, Fayetteville  
MAIN  
7

Received on: 09-29-06  
Inquiry : the University of  
Arkansas undergraduate  
research journal

**INQUIRY: The Undergraduate Research Journal of the UNIVERSITY OF ARKANSAS, FAYETTEVILLE**

**Volume 7 - 2006**

2 Editor's Foreword, Publication Board

**Section I: Arts and Humanities**

---

- 4 ARCHITECTURAL HISTORY AND THEORY: *Callie Williams*. Development of Gendered Space: The Archaic and Classical Greek Temple
- 10 ART: *Jennifer Pease*. Lucas van Leyden's *Dance Around the Golden Calf*: The Northern Triptych in the Age of the Reformation (Note: This work was recognized as one of the three best undergraduate research projects completed in 2005-06.)
- 18 FOREIGN LANGUAGES: *Lauran Elam*. *Beurs in the Hood*: Coming of Age in the *Banlieue*
- 29 MUSIC: *Ana María Alarcón Jiménez*. The Sounds of Japanese Noise: First Generation of Japanese Noise-Artists

**Section II: Social Sciences and Business**

---

- 37 ACCOUNTING: *Rikva Berman*. SEC Enforcement Actions: An Analysis of Economic Determinants of Restatements
- 44 ACCOUNTING: *Andrea Parker*. An Investigation of Changes in Contributions of State Lotteries to Education Over Time (Note: This work was recognized as one of the three best undergraduate research projects completed in 2005-06.)
- 52 FINANCE: *Elizabeth Dodson*. Semi-strong Form Market Hypothesis: Evidence from CNBC's Jim Cramer's *Mad Money* Stock Recommendations
- 59 INFORMATION SYSTEMS: *Brian Thanh Tran*. Biometric Security: Are Inexpensive Biometric Devices Reliable Enough to Gain Wide-Spread Security Usage?

**Section III: Sciences and Engineering**

---

- 66 BIOLOGICAL AND AGRICULTURAL ENGINEERING: *Jack Bourne* • Evaluation of Rice Hulls as a Lignocellulosic Substitute in Wood Plastic Composites (Note: This work was recognized as one of the three best undergraduate research projects completed in 2005-06.)
- 73 CHEMICAL ENGINEERING: *Thet N. Soe*. Ultrafiltration Membrane Process for Pyrogen Removal in the Preparation of Water for Injection (WFI)
- 81 COMPUTER SCIENCE AND COMPUTER ENGINEERING: *Kyle Neumeier*. Dynamic Composition Of Agent Grammars
- 93 INDUSTRIAL ENGINEERING: *Mauricio Carrasco*. A Study of the Impact of Prognostic Errors on System Performance

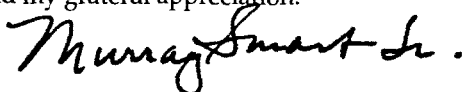
**Foreword:**

This journal, the seventh in an annual series, is a project of the Teaching Academy of the University of Arkansas and is testimony to the Academy's belief that a function of good teaching is to encourage good research and creative thinking on the part of the students.

This issue of *Inquiry* records the individual research exploration of twelve U of A student/faculty mentor pairs during the 2005/2006 academic year. The projects included here are drawn from disciplines from four of the six undergraduate colleges and schools—the School of Architecture, the J. William Fulbright College of Arts and Sciences, the Sam M. Walton College of Business Administration, and the College of Engineering—and are representative of the quality of research done by honor students in the various disciplines represented on campus. The breadth of subject matter included here is testimony to the commitment made throughout the university to honors study and research at the undergraduate as well as the graduate level. These twelve articles were chosen by *Inquiry's* publication board from nearly fifty papers submitted for consideration. (Publication economics controls the size of the journal: most of the papers received were of publishable quality.) The papers chosen vary in subject, in writing style, and in the manner in which they reference their research sources; but they are uniformly excellent in content. In most cases, the paper published herein is a digest of the student's larger research product. The intent of the journal is to record the depth and breadth of the scholarly activities of the university's best undergraduate students. I believe that it does this.

The editor and publication board of *Inquiry* depend on colleagues throughout the institution to serve as reviewers of material submitted for consideration. The range of disciplines at an institution like ours makes it impossible for the board to be able to evaluate all the papers submitted. We are grateful for the assistance that we receive from reviewers throughout the institution.

A personal note: This is the last issue of *Inquiry* for which I will serve as editor. The journal will continue as a project of the Teaching Academy, of course, but under new leadership. I have enjoyed my tenure as *Inquiry* editor and am proud of the quality of the student work published since the birth of the journal seven years ago. The journal as it exists today would not have been possible without the wholehearted support of the Provost, Robert Smith, the board and membership of the Teaching Academy, the *Inquiry* Publications Board, faculty who encouraged their students to submit their work to the journal for publication consideration, and to the students whose work we have published. To all these people I extend my grateful appreciation.



Murray Smart, Jr., Editor, University Professor of Architecture, Emeritus

---

### INQUIRY PUBLICATION BOARD, 2005-2006 ACADEMIC YEAR

Ethel Goodstein, Professor of Architecture; Douglas A. James, Professor of Biological Sciences; Donald M. Johnson, Professor of Agricultural and Extension Education; Thomas Kennedy, Professor of History; John Norwood, Associate Professor of Accounting; Curt Rom, Associate Professor of Horticulture and President of the Teaching Academy; Charles Rosenkrans, Associate Professor of Animal Science; Mary Jo Schneider, Professor of Anthropology; David A. Sloan, Associate Professor of History; Deborah Thomas, Professor of Accounting and President-elect of the Teaching Academy



## **SECTION I: ARTS AND HUMANITIES**

**ARCHITECTURAL HISTORY, ART HISTORY, MUSIC HISTORY,  
AND FOREIGN LANGUAGES**

## DEVELOPMENT OF GENDERED SPACE: THE ARCHAIC AND CLASSICAL GREEK TEMPLE

By Callie Williams  
Program in Architectural Studies

Faculty Mentor: Dr. Kim Sexton  
Department of Architecture

### Abstract:

*Throughout the ancient Greek world, temples marked the landscape as a sign of Greek civilization. Although Greek temples have been examined, described, and catalogued scientifically since archaeology came of age in the 18<sup>th</sup> century, the question of their cultural significance in their original Greek context has yet to be fully answered. Many twentieth-century interpreters tended to yoke the history of Greek temples to narratives of modernism, resulting in anachronistic conclusions. Current trends in architectural history have begun to test other interpretative strategies, such as the interrelationship between architecture and the emergence of Greek philosophy. This essay explores the possibilities of a deeper reading of these architecturally stunning spaces through both current and classical theories of gendered space. Despite thorough Freudian examinations of gendered space as it relates to the body and cultural interactions in modern domestic architecture by sociologists such as Steven Pile, culturally significant buildings such as the Parthenon have remained unexamined by gender theorists. Furthermore, the recent scholarship in Classical body theory is bringing to light more and more data on the Greeks' own attitudes toward gender as it was inscribed on the body—unfiltered by the Freudian lens. My investigation centers on the development of peristyles and cellas of signature temples throughout the Greek realm, demonstrating how culturally inscribed gender ideals influenced, and even shaped the evolution of the temple form from the Archaic through the Classical period of Greece.*

### The paper:

Culturally ingrained notions of gender shape the spaces of daily life. Ancient Greece was no exception; in fact, given the immense influence of classical Greek culture on the West, Greek gender practices merit close attention. As the people of Ancient and Classical Greece shaped the physical spaces of their daily lives, the idea of gendered space evolved; thus becoming a more and more significant player in architectural design. One can

view specific forms of architectural space through the Greek ideals of gender hierarchy.. Initially, the function and occupancy of certain typical spaces created structures with inherent conditions of internal and external space. Greek ideals of lived space – i.e., space lived directly through associated images and symbols – would have influenced the design of both domestic (e.g., the Greek house or *oikos*) and public venues (e.g., the agora).<sup>1</sup> The Greek *oikos*, the protective abode of the ideal Greek woman, hidden away from public interaction, became the archetype of what is now known as feminine space. However, the gendered “femininity” of its space does not reside in the presence of female bodies dwelling therein. Instead, it is the main characteristics of hidden or enclosed space with often very dim lighting that conveyed a sense of fixity and security, a much sought after characteristic of a residential dwelling, which genders this space feminine.<sup>2</sup> In contrast, the space of the Greek agora is masculine in that it is characterized by an unremitting sensation of openness and mobility. These qualities combine to create architectural forms that foster the idea of challenge and defiance. This idea is readily evident within the Greek culture of public debate or rhetoric which is reserved exclusively for male citizens. These spatial readings of the feminine as dark, fixed, and secure, and the masculine as open, expansive, and mobile are essentially Freudian, as sociologist Steve Pile has demonstrated.<sup>3</sup>

This essay seeks to augment these mid 20<sup>th</sup>-century readings through the use of Greek cultural definitions of the feminine and the masculine as espoused in primary sources of the time.<sup>4</sup> Unlike 20<sup>th</sup>-century Western culture, Greco-Roman texts on natural science and medicine did not see gender as biologically related to reproducing. Instead, they saw gender as a spectrum of traits running from the extremely feminine to the extremely masculine. Where an individual was located on the spectrum depended not on his or her biological sex, but on his or her personal characteristics. On this spectrum, the Greeks identified the feminine body as cold, wet, leaky, and heavy, whereas the male body was hot, dry, and airy. Furthermore, rather than examine the traditionally “feminine” domestic space versus the “masculine” public space, this paper proposes to explain the

evolution of the Greek temple from pre-Archaic Greece to the Classical period in terms of gender. In so doing, it maps the temple from an original feminine space, through a period of mixed gendered space, to reveal the Greek temple at the pinnacle of its Classical development (e.g. the Athenian Parthenon) as intrinsically hermaphroditic.

According to most contemporary and ancient authors, the development of this most recognizable of ancient icons, the Greek temple, evolved along a strictly linear constructional ethos that moved from modes of wood construction translated into stone to the later translation of democratic and civic ideals into structural ethics. Vitruvius extolled this idea in his *Ten Books on Architecture*:

"And thus each element preserves its proper place, type and order. Drawing from these elements and from the art of carpentry and applying them to the construction of sacred dwellings in stone and marble, craftsmen imitated these arrangements in their sculptures and agreed that these inventions ought to be adopted. ... And thus the covered section of the joists in Doric works began to take on the arrangement of the triglyphs and, between the joists, the metopes." (4.2.2)<sup>5</sup>

In the 20<sup>th</sup> century, archaeologist and architectural historian William Bell Dinsmoor attributed the first stages of development to advances in constructional technologies and methods, but he also suggested that there must be something more to the collective Greek strategy of temple building:

"Mere advances in construction do not account for the development of the shrine; it is of the aspiration of humanity toward something fulfilling their ideal of a house of God that the Greek temples speak. Building better than they wot of, one generation joined hands with another in rearing these most splendid fabrics of in-dwelling divinity."

These ideas are clearly deeply imbued with contemporary architectural discourses on the idea of form following function or functionalism. Additionally, they reflect, anachronistically, the organization of architectural production of the 20<sup>th</sup>-century which includes the professionalization of architecture and the organization of construction into trades, neither of which existed in ancient Greece. While Dinsmoor calls for alternative interpretations, he does not elaborate further. I propose that the ancient Greeks looked to their own subconscious ideas of gender and social structure. Indeed, the idea of gendered space and its impact on the development of the Greek temple allows a more in depth reading of all aspects of a purposefully created space rather than merely an alternative way to ascribe meanings to a space.

The Greek temple, a public space of personal and collective devotion, epitomizes the idea of gendered space and its changing conception throughout ancient and classical Greece.<sup>6</sup> During its

humble beginnings in the so-called "Dark Age" of Grecian history circa 1100-800 B.C.E., the typical temple resembled a small house, or earthly abode, for the patron deity (Figure 1). What would later become the cella, or inner sanctuary of the classical and archaic temple, was originally a small and simple wooden or mud brick hut-like structure with a triangular roof. Usually, columns supported a small eave over a doorway on one end of the structure. Inside this structure a cult statue would be placed.<sup>7</sup> This structure is a clear embodiment of the ideals of feminine space; dark, hidden interior, protected by load bearing walls and a great sense of immobility due to the weight and closeness of the structure and the internal space the structure created. The enclosed area of veneration of the deity, augmented by low lighting due to the overhanging roof structure and eaves, amplifies the experience of the mysteries of the god or goddess within. Also, interior columns, placed to support the roof structure led to a dissection of the internal space, keeping the whole interior from being viewed at one time. These intentional design insertions all give evidence of a feminine space; identified by a strong sense of security and mystery.

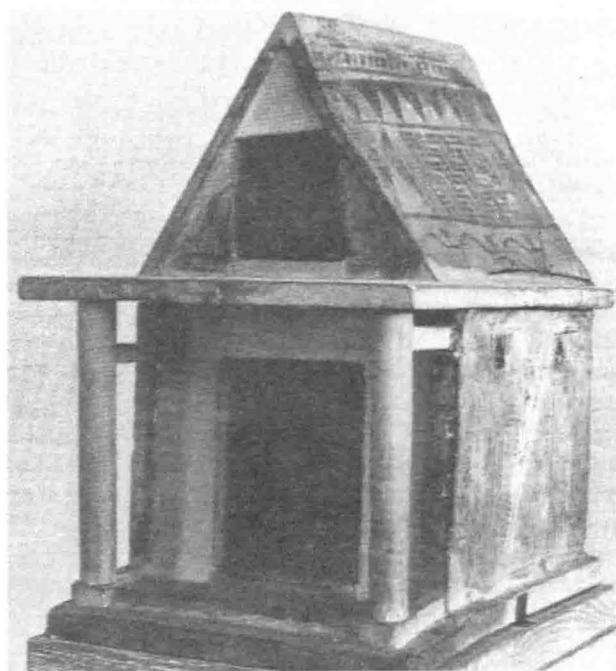


Figure 1 Greek shrine, votive clay model, from Argos (Greece), eighth century B.C. (Kostof, 117)

During the Archaic period circa 600-480 B.C.E., the feminine space of the cella, dedicated to veneration of the deity, was further elaborated and further feminized. The original temple to Hera, the queen of the gods, at Samos is a clear illustration of the way in which the interior of the cella was extended axially as early as 800 B.C.E, in this case to nearly 33 meters in length, creating a hecatompedon or temple that is 100 feet in length. Also, within Hera I at Samos, the interior set of wooden columns which had been placed irregularly in the pre-

archaic temple for roof support now fell in line directly in front of the cult image.<sup>8</sup> This allowed the cella to become a passage down which a believer would walk without full view of the goddess until reaching the very end of the space (Figure 2). Therefore, the hidden or feminine space of veneration at Hera I at Samos was elaborated on a more monumental scale through simple design regularizations.

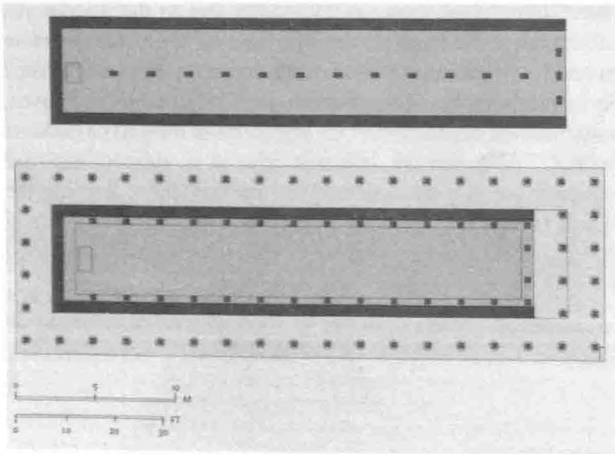


Figure 2 Hera I, c. 800 B.C., and Hera II, c. 650 B.C., at Samos, Greece (Stierlin)

Nearly fifty years after the completion of Hera I at Samos, gendered masculine space made its first timid appearance in Greek temple architecture. This came in the form of the encircling peristyle of wood columns at Hera I. This radical new development in temple construction was more than a simple awning of wood. It produced an entirely new envelope for what had once been a flat façade. Now, masculine ideas of control began to exert themselves over what was to become the inner sanctuary. This new masculine space, as defined earlier, promoted an idea of mobility by offering the space up as an analog to the space around it or as Kostof would say a mid-space object. (170) Although conceived originally as an impermanent wooden structure, the idea of the surrounding peristyle would eventually become the iconographic symbol of Grecian architecture.<sup>9</sup>

A more intact and well-known temple to Hera, located at Paestum, Italy was constructed circa 530 B.C.E.. This temple illustrates the archaic temple at its height according to 20<sup>th</sup> century critics. This icon is one of the first to combine feminine and masculine gendered space in the sacred enclosure of a Grecian temple through the use of permanent materials.<sup>10</sup> The interior cella of this temple is still much like the original temple dedicated to Hera at Samos with a centralized row of columns (Figure 3). Although the temple form is now solely made of stone, the aspects of its construction are still the same. The space created by the peristyle here at Paestum, although masculine at

first glance, still displays some aspects of femininity once its structure is analyzed. The columns that make up the peristyle show a pronounced entasis or a swelling at the column's center so that the column's diameter is wider at its center than at its mid-sized base and smaller top. This entasis creates a feeling of weight and support as the structure is viewed as a whole (Figure 4). This focal component makes the structure seem heavy and tied to the ground, opposing the sense of mobility that should accompany masculine space. Although it seems an attempt was made, the later temples of Greece create the masculine envelope much more effectively by combining the femininity of the column with the idea of masculine penetration of space, giving weight to the argument that the constant search for perfection in temple design was in reality a hegemonic struggle between gendered spaces.

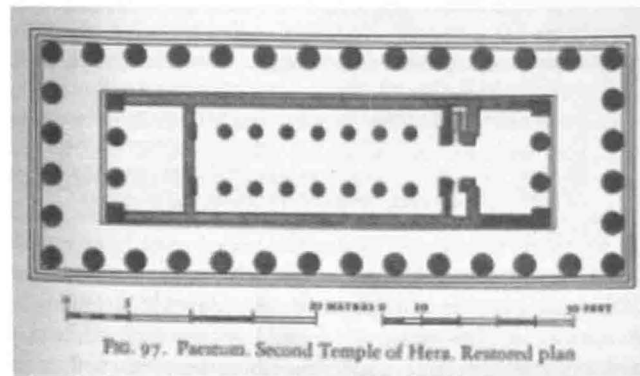


Figure 3 Plan of Hera II at Paestum (Scully, FIG. 97)

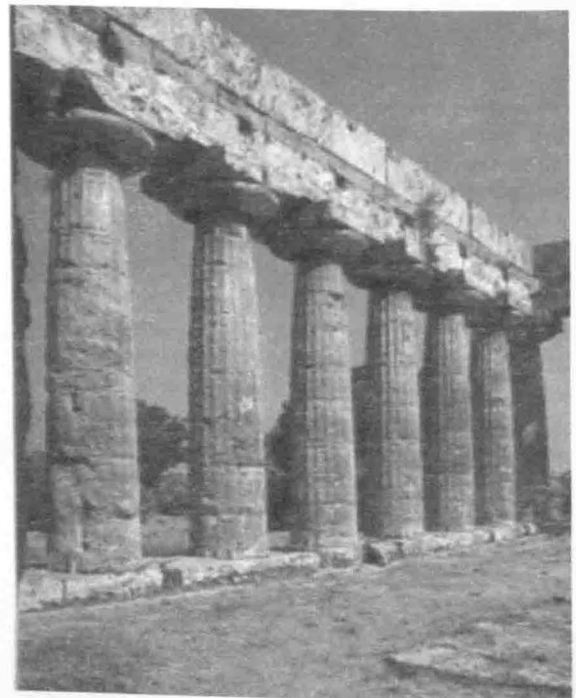


Figure 4 Hera I at Paestum (entasis in peristyle) (Stierlin)

The idea of gendered spatial formation within Greek temples was not just limited to temples of female deities. Unlike in today's society, biological sex and gender were not as intrinsically linked in the classical era of Grecian culture.<sup>11</sup> Femininity was not seen as just having female sexual organs, but as a mind set or a way of life separate from reproductive organs. Culturally determined gender ideals are inscribed onto bodies, buildings, and things animate and inanimate. This idea is readily evident when switching to temples of male deities. One example is the Temple of Poseidon, a Greek god of the sea, located in Sunium overlooking the Aegean Sea. This Doric temple, constructed in the 5<sup>th</sup> century B.C.E., measured 100 feet in length including a peristyle that enclosed a central cella (Figure 5). Once completed, the entire structure was composed of marble. It is one of the best examples of how Greek architecture used the surrounding landscape to imbue a building with a sense of a true link to the natural world.<sup>12</sup> This sense of presence, afforded by the peristyle's challenge to the natural world was a totally masculine ideal. The repetitive light and shade of the columns contrasted starkly with the flat plane of the sea beyond, offering a dynamic feeling of challenge and defiance capitalizing on the Greek understanding of the impact of the natural landscape (Figure 5). Also, the cella was constructed without the interior set of columns, allowing the space to become even more stark and cave-like than its predecessors. This cave-like interior harkens back to the idea of a feminine space as an analogue to a "mother earth" condition. Thus, the ideal temple creates an enclosed interior which is nothing like the expansive nature of the god of oceans and seas that it resides within, demonstrating that the form of the cella was a continuous act of historical design in line with other temples and not specific to the sex of the deity



Figure 5 Temple of Poseidon, Sunium, Greece (columns against Aegean Sea) (Stierlin)

The epitome of classical Greek temple design is most widely thought of as the Parthenon, the renowned temple to Athena Parthenos, the impenetrable maiden goddess. An icon of the Greek nation, the Parthenon is the crowning temple of the

Acropolis in Athens, Greece. A huge Doric temple in its own right, the Parthenon dominates the city of Athens from its high position on the ancient acropolis. The columns creating the peristyle respond to the challenge of the environment through their large scale and logic-defying entasis (Figure 6). According to Indra Kaxis McEwen the peripteral elements of the classical temple became like sails or *ptera*, a Greek word meaning wing.<sup>13</sup> The change from a heavy and deeply grounded reading of space, structure, and peristyle such as at Hera I at Samos and Hera I at Paestum to a space filled with tension between the reality of load and support and a conveyed sense of uplifting and lightness allowed the peristyle to become a more masculine space, embodying all of the fundamental characteristics of challenge and defiance as well as the airiness of the model elite Greek male body. The interior of the Parthenon however retains its fundamental femininity. Indeed, in its proportions, its frieze, and its double row of porch columns, the "feminine" Ionic order is introduced into Doric temple architecture. Even Vitruvius imbues the creation of the Ionic order with a feminine identity:

"...the Ionians built a temple to Diana; seeking a new type of appearance, they applied the same ratio based on footprints to a woman's slenderness, and began making the diameter of the columns measure one-eighth their height, so that their appearance would be more lofty. Instead of a shoe, they put a spira underneath as a base, and for the capital, as if for hair, they draped volutes on either side like curled locks..." (4.1.7)

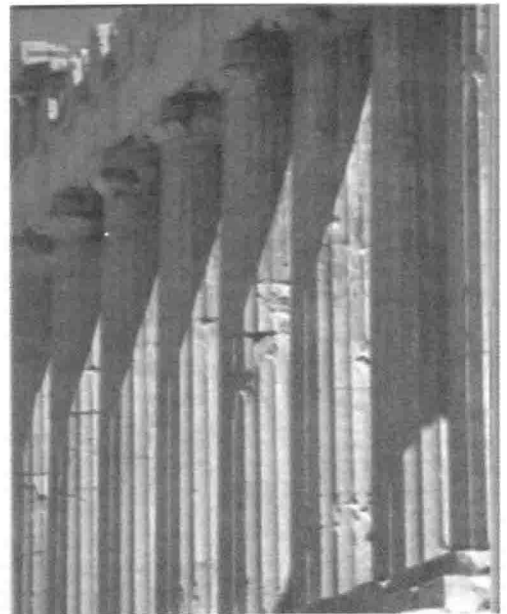


Figure 6 Temple of Athena Parthenon, Athens, Greece (Stierlin)

Furthermore, the interior engaged columns allow the space to become subdivided, but without blocking the view of the grandiose gold and ivory cult image of Athena that once dwelled within the cella (Figure 7). Although the view of the cult image



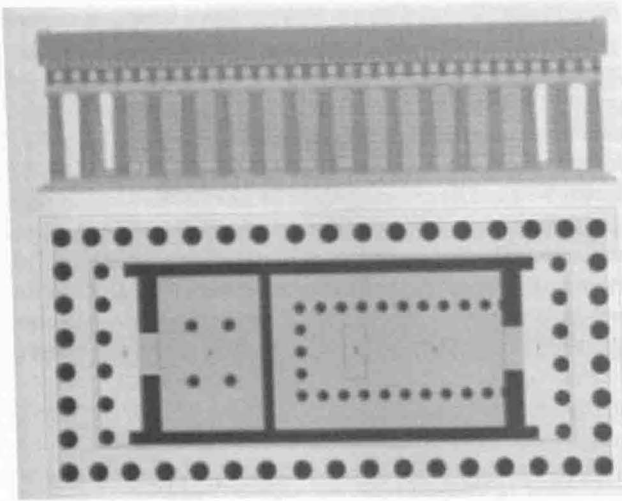


Figure 7 Temple to Athena Parthenos, Athens, Greece (Stierlin)

is no longer blocked, the interpenetration of space and column within the cella allows the space within the structure to seem dense and heavy obscuring the plainer wall with deep shadow.

The transition from the Archaic to the Classical era in Greek temple design created an iconic building type that has lasted throughout the ages. The exterior envelope created by the peristyle of columns produced a space which seemed to hermaphroditically and protectively hug the interior cella space of veneration of the god or goddess. Although initially feminine in structure and feel, the peristyle became a more masculine space as the search for perfection generated an envelope that protects the inner space of the cella structure while still celebrating its own separate identity. The internal cella was in every aspect a feminine space, perfect for representing the earthly presence of a deity. This was essential in all structures from the Archaic to the Classical era, as the interior of Grecian temple remained a space retained for symbolizing the fixity and continuity of an unearthly being. This idea was even represented in the archaic temple through the chaining of the cult figure to keep it from escaping the cella space<sup>14</sup>. Each of these specifically gendered spaces allows a fuller, more dynamic reading of cultural advancement in Grecian design; from the Archaic to the Classical era and beyond.

#### Works Cited:

- Anderson, Williams J. and Williams Bell Dinsmoor. *The Architecture of Ancient Greece: An Account of Its Historic Development*. London: B. T. Batsford Ltd., 1978.
- Cole, Susan Guettel. *Landscapes, Gender, and Ritual Space: The Ancient Greek Experience*. Berkeley: University of California Press, 2004.
- Gonzales, Matthew. "The Binding of Ares in Myth and Cult: A Re-assessment." *American Philological Association*. 31 October, 2005. <http://www.apaclassics.org/AnnualMeeting/05mtg/abstracts/GONZALES.html>
- Kostof, Spiro. *A History of Architecture*. New York: Oxford University Press, 1995.
- Martin, Dale. *The Corinthian Body*. New Haven: Yale University Press, 1995. pp 3-37.

McEwen, Indra Kagis. *Socrates' Ancestor: An Essay on Architectural Beginnings*. Cambridge: The MIT Press. 1993. pp 79-120.

Pile, Steve. *The Body and the City: Psychoanalysis, Space and Subjectivity*. New York: Routledge, 1996. pp 210-237.

Scully, Vincent. *The Earth, the Temple, and the Gods: Greek Sacred Architecture*. New Haven: Yale University Press. 1979.

Stierlin, Henri. *Greece from Mycenae to the Parthenon*. New York: Taschen, 1997.

Vitruvius. *Ten Books on Architecture*. Ingrid D. Rowland and Thomas Noble Howe. New York: Cambridge University Press, 2002.

#### Endnotes:

<sup>1</sup> This is a reference to Henri Lefebvre's tripartite theory of space and special meaning in which lived or representational space is the "space as directly lived through its associated images and symbols, and hence the space of 'inhabitants' and 'users'." (39).

<sup>2</sup> In *Landscapes, Gender, and Ritual Space*, Susan Guettel Cole explains more thoroughly the idea of home as related to femininity through the idea of gendered language. "For the Greeks, the earth (Ge or Gaia) was mother of all—not only feminine in grammatical gender but female in function: a source of nourishment and birth mother of the gods." Also, Cole illustrates the Greek idea of female and male space. "The undifferentiated earth was mother, but the land bounded by recognizable borders and a territory to be defended was a fatherland." (1-2)

<sup>3</sup> Steve Pile's discussion and exploration into the different ideas of Freudian gender as related to space and interpretation of space confronts the uses of the horizontal plane, nature and home to represent the feminine. This can be interpreted as fixed and restful ideal. Pile also discussed the use of the vertical plane and symbols of challenge and defiance as forms of male gendered identity. Although these ideas are inherently phallic, they can also be interpreted as symbols of power and mobility when compared to the horizontal plane and restful nature of the female symbols. (215–217, 221, 223)

<sup>4</sup> Dale Martin discusses at length his interpretation of classical source material, especially medical treatises of the Greco-Roman world, in his work *The Corinthian Body*. (3-37)

<sup>5</sup> This citing refers to Book 4, Chapter 2, Section 2 of the Vitruvius translation by Rowland and Howe. All other citations of Vitruvius follow this format.

<sup>6</sup> In the third book of *Ten Book on Architecture* written by Vitruvius, the intentional design of temple structures as relating to personal devotion is clearly evident: "When matrons climb the steps of the pycnostyle temple to make their prayers, they cannot walk arm in arm through the intercolumniation; they must go in single file. Furthermore, the view of the doorways is blocked by the close placement of the columns, and the cult statues themselves are half hidden; also because of the restricted space, movement round the temple portico is hindered." (3.3.2)

<sup>7</sup> Cult statues, their placement and the relative spatial practices surrounding them are discussed by Kostof in *A History of Architecture: Settings and Rituals*. (117-118); McEwen, 4-6, 56, 115-16.

<sup>8</sup> Placement of the cult statue of Hera is determined according to diagrams of Hera I at Samos included in Stierlin's *Greece from Mycenae to the Parthenon*. (72, 104)

<sup>9</sup> McEwen cites Vincent Scully on the mid-space element theory: "Vincent Scully says that the peristyle was intended "to articulate, penetrate, and extend the exterior envelope of the building so that it should become a true mid-space element, at once bounded and boundless ... setting up with its columns ... a regular standard of measure whereby distant horizons could be grasped." Although this premise is refuted in McEwen essay, I proposed that this reading is correct, yet only as it applies to the tensions and challenges established as a characteristic of male gendered space. (101)

<sup>10</sup> Hera I at Paestum is an early and well-preserved Doric temple completely built of stone; a characteristic that would come to symbolize the Greek temple. The earliest known entirely stone Doric temple was the Temple of Artemis at Corfu, built circa 580 B.C.E.

<sup>11</sup> Refer to Dale Martin's interpretation of classical conceptions of

gender and the body in *The Corinthian Body*.

<sup>12</sup> In Scully's *The Earth, the Temple and the Gods: Greek Sacred Architecture*, the Temple of Poseidon at Sunium is described as follows: "The temple of Poseidon is a double boundary, a kind of fence, first against the empty sky and the sea and then against the land.... As the worshipper mounted toward the colonnade, the stoa opened a perspective to the right toward the island of Patroclus and the horned promontory opposite it, blocked off the land behind it and directed the eye, as did the columns of the temple itself, toward the closely bounded sea view to the west." (162-163)

<sup>13</sup> McEwen authored *An Essay on Architectural Beginnings* which traces the development of the phonetic culture of the Greek temple.

<sup>14</sup> This point is briefly discussed in McEwen's *Socrates' Ancestor*: "The Heraion at Samos was the home of the cult image of Hera, a xoanon that was kept chained up in order to reveal the fearful dynamism of its divine life. The first mud brick Heraion, with its interior hearth place, was a house writ large, and it is in the nature of houses with their hearths to be fixed, to be anchored to the soil." (101) "

#### Editor's note:

This paper won the 2006 Pella Student Writing Award--a juried competition open to all students in the school of Architecture.

#### Faculty comments:

Dr. Kim Sexton makes a cogent case for the excellence of her student's work. She writes,

Ms. Williams wrote this provocative essay as a research project in fulfillment of the requirements for the Honors section of History of Architecture I in fall semester 2005. In spring 2006, enrolled in my interdisciplinary Honors Colloquium, *Medieval Bodies / Medieval Spaces*, Ms. Williams proceeded to refine her argument through the addition of Greco-Roman body theory introduced in this seminar. She revised her paper for submission on her own, independently of the Honors Colloquium course requirements. Thus, this essay is the product of one-year's research. I have advised Ms. Williams and reviewed her progress throughout this year.

I would like to highlight five qualities of Ms. Williams' essay which place it among the best research produced by undergraduates (and by graduate students for that matter):

Ms. Williams' account of the evolution of Greek temple form through gender theory is utterly original research. No scholar has yet inscribed the Classical sex-gender pyramid on Greek temple architecture.

Ms. Williams' agile handling of sophisticated theoretical paradigms - Freudian spatial analysis and body theory - structures the analyses throughout her essay. Here, theory is no mere introductory distraction or flourish.

In the formal analysis of architectural remains - still the key interpretative tool of architectural historians - this essay shows Ms. Williams to be not only competent, but progressive and original.

Ms. Williams clearly defines the position of her research within the historiography of the Greek temple.

Finally, what truly sets this essay apart is Ms. Williams' avoidance of the pitfalls of essentialist reasoning that ensnares many undergraduate researchers. She works deftly with a sliding spectrum of traits rather than giving undue credence to artificially constructed categories such as masculine vs. feminine, sacred vs. secular, etc.

Ms. Williams' submission, "Development of Gendered Space: The Archaic and Classical Greek Temple," is a model essay of undergraduate research on a topic of fundamental importance - gender and architectural design - both for the discipline and for society at large.



# LUCAS VAN LEYDEN'S *DANCE AROUND THE GOLDEN CALF*: THE NORTHERN TRIPTYCH IN THE AGE OF THE REFORMATION

By Jennifer R. Pease  
Department of Art

Faculty Mentor: Lynn F. Jacobs  
Department of Art



Figure 1. Lucas van Leyden. *Dance Around the Golden Calf*, c. 1529-30. Rijksmuseum, Amsterdam. (© Rijksmuseum)

## Abstract:

*This paper focuses on Lucas van Leyden's 1530 triptych Dance Around the Golden Calf. It is shown through a stylistic and iconographic analysis of the painting that it is a piece reflective of the tensions and upheaval prevalent during the Reformation-era society in which it was produced. This was a*

*time when Northern European artists such as Lucas began an increase in secularism in their work, demonstrating the dawn of a new era in the church. Lucas effectively uses visual contradiction and psychological implication to respond artistically to the tumultuous age of the Reformation.*



Figure 2. Lucas van leyden. *Dance Around the Golden Calf*. Central panel.

### Introduction:

Lucas van Leyden's *Dance Around the Golden Calf* (Figures 1 & 2) is a sixteenth century triptych which reflects the tension of the transitional era of the Reformation in its style and iconography. Lucas was creating this painting during the dawning of a new age in the church. This tumultuous time sparked much debate about the role of the church in daily life, and people began discussing revolutionary religious ideas. One of the primary sources of debate centered on the importance of the taking of Sacraments, or communion, in the church. Lucas remarks through *Dance* on the contradictions and topics prevalent during this time. This paper seeks to demonstrate how Lucas's painting embodies a multifaceted synthesis of Italian Renaissance and Dutch Netherlands influence to create a simultaneously cohesive and contradictory work of art which reflects the temperament of the Reformation Age. It will be shown that *Dance Around the Golden Calf* is a "composite, highly mannered art, as realistic in details as unrealistic in general effect; an art that reflected the spiritual ferment of the age, its skeptical detachment, its nervous tension, and its intellectualistic bias" (Leymarie 51).

### Style:

Lucas's treatment of style in *Dance* reflects the tensions and dualities of the Reformation. The color in Lucas's *Dance* shows the influence of Northern masters before him, and also exhibits a personal mastery of the element of balance in his work. Essentially, Lucas blends Northern elements with new Italianate ones. The use of bright reds is characteristic of the North. The concept of complete color balance and harmony reflects the ideals of the Italian Renaissance. The color red is repeated rhythmically throughout the composition of *Dance*, drawing the eye in cyclical patterns throughout the work. It is notable that these undulating patterns are also repeated in buttocks, jugs, baskets, and clouds, as well as the dwellings of the Israelites in the background of the piece. Their significance will be discussed later in the paper. Visually, the red color of *Dance* acts to unify these repeating circular patterns, creating harmony and movement. A similar use of red as a device for movement may be seen in many earlier Northern works, an example of which is Jan Van Eyck's *Ghent Altarpiece*. Lucas also used a simpler method of red as a vehicle for circular movement in his 1515 piece, the *Card Players* (Figure 3). The color scheme Lucas chose for *Dance* is largely subdued other than the rhythmic red. The palate of earth tones, including a range of cool browns, yellows, greens and blues serve as a balancing mechanism to the warm, attention grabbing red. Lucas juxtaposes these warm and cool hues to achieve a complex tonal balance within the work (Smith, New Appraisal, 39). Compositionally, the depth of hue in the foreground contrasted with the lighter handling of color in the middle ground is complimentary to Lucas's innovative inverted format, discussed later in further detail.

We once again see the Italian Renaissance influence in Lucas's treatment of the human figure and enveloping drapery in *Dance*. The skill and neatness with which the drapery is rendered evidences Lucas's background as an engraver (Stechow 35). One example of Lucas's ability to render drapery realistically may be found in his *Temptation of St. Anthony* (Figure 4). The drapery folds in *Dance* have the same realistic modeling over thick, sturdy forms that are apparent in the *St. Anthony* engraving. Italianate influence is manifest through Lucas's sculptural treatment of the form, reminiscent of Donatello's bulky, monumental body types. The corporealism found in Lucas's treatment of the figure in *Dance* reinforces the piece as a genre subject dealing with morality issues. The Israelites are engaged in self-gratification, and the voluptuous forms with clinging drapery reinforce the sensual nature of their actions (to be discussed further in the second section of this paper). Furthermore, throughout *Dance* there are many figures which embody Italian prototypes of the figure, for example the back view of the female in the centralized portion of the left panel. This figure in particular may also show Albrecht Durer's influence on Lucas's style, for it is quite similar to Durer's *Nude From Behind* (Snyder 322), a piece which was ultimately influence by the Italian



Figure 5. (Above) Raphael. *Madonna and Child and Young St. John (The Alba Madonna)*. National Gallery of Art, Washington, D.C. (©Alinari / Art Resource, New York)

Figure 3. (Left) Lucas van Leyden. *The Card Players*. The National Gallery of Art, Washington, D.C.



Figure 4. Lucas van Leyden. *The Temptation of St. Anthony*, 1509 (engraving).



Figure 6. Lucas van Leyden. *David Playing the Harp Before Saul*, c. 1508 (engraving).

Renaissance. The female figure in the center panel of *Dance* has another Italian prototype, Raphael's *Alba Madonna* (Figure 5). Furthermore, the Italian influence on body types and posturing is apparent in the female figure in the centralized portion of the right panel. This female form is reminiscent of the one observed in Botticelli's *Birth of Venus*. Lucas's voluptuous form stands in a graceful sway, the contrapposto stance highlighted by the treatment of drapery, noted especially on the right leg. The prototypes Lucas uses throughout *Dance* reinforces the artist's references to the Italianate tradition.

Lucas's treatment of faces in *Dance* is another stylistic element which reinforces the contradictory and transitional nature of the piece. In *Dutch and Flemish Painters*, C.V. Mander observed that Lucas possessed great skill in depicting emotions, as seen in Lucas's earlier work *King Saul and David* (56), (Figure 6). This scholar's observation made me realize that the opposite phenomenon is at work in Lucas's *Dance*. All of the faces in *Dance* are depicted in a neutral fashion. Some stare straight out at the viewer, but the look in their eyes is expressionless. No mouths smile, no faces grimace in this piece. Some faces, such as the centralized female in the right panel are entirely homogenous, the features completely undefined. This tactic was used by Lucas to spark a dialogue between the painting and the viewer, provoking the viewer to draw their own moralistic conclusions about the piece. Another observation concerning the facial qualities in *Dance* is in reference to a woman's face which appears in the lower right foreground of the center panel. The wide face, shifty eyes, and flat headpiece recall in my mind

a face in the right panel of Bosch's *Garden of Earthly Delights*. This similarity shows possibility that Lucas was influenced by his contemporary, and may also act as a sly allusion to Hell when connecting the face in *Dance* with the earlier Bosch work.

It is commonly accepted by scholars that Lucas's *Dance* was executed in 1530, which was late in his career. Lucas was a fabled child prodigy who had a natural proficiency for engraving, creating accomplished work as early as age nine (Mander 48). Certain visual elements in Lucas's early work, such as hands gripping sticks and staffs, are found in the artist's early works such as *Susana and the Elders* and the *Milkmaid* (Figure 7). These visual elements resurface in Lucas's *Dance* in a more sophisticated manner, the significance of which will be discussed in respects to iconography. The shape of the *Dance* triptych is also an indication that Lucas executed this piece later in his career. The bell shaped format began appearing in much sixteenth century art, and its break from the traditional rectangular format was a necessary innovation of artists striving to capture the public's attention in a shifting market. Lucas experimented with a similar shape in his 1526 *Last Judgment* triptych prior to his use of this shape in *Dance*. This is a shape that is not manifest in Lucas's earlier work, the majority of which are engravings.

There is a conflict between Lucas's treatment of the exterior of *Dance* versus the interior which reiterates the tension prominent in the piece. In assessing the exterior to the piece, greater explanation may be found as to the intent for the location of the work. The outside of the triptych is rendered simply in green and



Figure 7. Lucas van Leyden.  
The Milkmaid, 1510 (engraving)



red imitation marble, devoid of subject matter (Smith 107). Lucas's failure to make the exterior more engaging is probably reflective of the notion that *Dance* was a secular piece, intended for personal home display rather than church altar function. The shift in Lucas's focus from the primary biblical story to the revelry of the Israelites may have also influenced his decision to not include the major figures from the story on the outside of the piece, as was the trend with previous Northern work. This subversion of the biblical story in preference of the emphasis on the Israelites' revelry is what makes this piece describable as an 'inverted composition'.

By using an inverted composition in *Dance*, Lucas creates a scenario in which the biblical is subverted by the secular. This brings the licentious behavior of the Israelites *rather* than the role of Moses to the forefront. This makes the piece focused on rebellion, rather than obedience, to God. This introduces an element of tension into the work which solidifies it as a commentary on the conflicted age in which it was produced. The viewer is not meant to simply view an illustrative tale of Moses descending from Mt. Sinai with the Ten Commandments. Rather, the viewer is accosted with the drunken revelry of the Israelites, their sensuous bodies filling the three triptych panels in the fore and middle grounds. We must strain to observe the biblical focus of the story, Moses, in the background. This inversion by Lucas shows how Northern Europe was shifting out of the middle ages towards the modern era. People were considering the role of the religious image in a new way. Rather than absorb a story illustrated straight from biblical text, they were thinking about the implications and deeper meanings behind the text, forming new opinions and ideas concerning religion.

The insignificance of the exterior of *Dance* is juxtaposed against the active, highly symbolic interior, and serves to heighten the sense of displacement and detachment that Lucas creates with his lack of facial expression. This is to say that Lucas creates a sense of incongruity in *Dance* by creating an energetic, complex composition which is full of active and emotive themes yet generalizes rather than humanizes his subjects. One would expect a lively, full composition to portray emotive participants, but rather they appear detached from their activities.

Lucas introduces further visual tension in *Dance* through his use of the three paneled triptych format. The artist uses this format to both unify and separate the visual presence of *Dance*. Division is apparent in that figures are cut off at the edges of the panels, for instance a face may be seen in the left portion of the right wing, but their body does not continue on into the next panel. However, the movement and momentum of the figures carries the eye fluidly throughout the piece. The continuous plane of the ground as well as the treatment of the skyline solidify the piece as simultaneously unified and divided. On the interior of *Dance*, visual unity is created between the three panels by the rhythmic red and repetition of cyclical patterning throughout. I

believe this cyclical patterning throughout the composition became a strong influence in much later work, an example of which is Matisse's *Dance II*.

One manner in which Lucas prevents the incongruities of *Dance* from creating an illogically conflicted piece is through repetition of motif. There is a repetition of trios in *Dance*, which unifies the piece visually and logically, and fits into the use of a triptych format. The female that creates the central axis of each panel occurs three times. Moses appears in the painting three times, the most difficult of which to recognize is the tiny Moses on the cliff of Mt. Sinai in the background of the center panel. In the right panel, the careful observer will note that besides the figure recognized as Moses, there are two other bearded men in close proximity. These Moses look alikes form a triangle in the right panel's middle ground. The pronounced spatial quality dividing the painting into definite foreground, middle ground, and background furthers the repetition of trios found in *Dance*. The possible meaning behind the unifying device of trios in *Dance* will be explored in the iconography portion of this endeavor. Further visual unity among the three panels of *Dance* is created by the expansive sky which flows through the piece. At first glance, the landscape seemed to reflect the same continuity, but upon further inspection, it seems the landscape contains both disparate and harmonious elements. This reiterates the earlier notion that *Dance* is simultaneously divided by the triptych format and incongruous faces, yet unified by the color and composition. On one hand, the flowing, painterly manner of the landscape is fluid from panel to panel. However, Lucas's decision to cut off trees, Mt. Sinai, and some of the camp of the Israelites introduces an element of tension into the piece. For example, the tree in the right portion of the center panel is severed abruptly, for if it continued it would actually be in front of the building directly adjacent to it in the right panel. Because the tree is cut off, a visual border is created at the right edge of the center panel which serves to contain the expansive space. The disparity caused by Lucas's decision to sever things at the panel parameters does not end with the landscape. The largest element of visual separation in Lucas's *Dance* is manifest by his repeated cutting off of the human forms from panel to panel. This is a unique decision by the artist, for in his earlier work *Last Judgment* the forms continue fluidly from panel to panel, an example of which may be found in the kneeling figure who carries over between the borders of the center and right panels. Why would Lucas disrupt the harmony created by color and form by repeatedly chopping off bodies at the panel edges? It is possible the artist was making a statement on the growing schism in the church. He may also have been alluding to the biblical wrath of God in Exodus 32:30, in which God tells Moses he will 'blot out' those who have sinned against him. Ultimately, Lucas's combination of harmonious and disparate visual elements in *Dance* results in an overall contradictory piece, the nature of which may have influenced the later Baroque style.

## Iconography:

Iconographically, Lucas's *Moses and the Worship of the Golden Calf* contains a complex layering of erotic, theological, and philosophical implications that show the piece to be a harbinger of a transient age. The dawn of the Reformation was a tumultuous time of collision and upheaval. Lucas was a member of the nobility of Leyden, married to a woman whose family was active in 'The Forty', which was the governing council at the time (Silver 406). This close contact with the ruling class kept the artist centered in the public eye, thus creating an importance for him to not be perceived as a heretic. Thus, Lucas was a Catholic who was compelled to incorporate Protestant ideas into his work (Silver 409). By using this combination of elements in *Dance*, Lucas creates a piece intended to act as a commentary of the times, as well as a catalyst towards independent thought. The centralized females in each panel, the repetition of stick and staff forms, and the image of wine and vessels are some of the primary images in *Dance* which contain deeper meaning. The following seeks to unravel some of the complex iconography of Lucas's *Dance*.

The erotic implications of *Dance* take a front seat to the biblical story for which the piece is titled. As discussed previously, the facial expressions, or lack thereof, in *Dance* are indicative of a desire to provoke moral pondering in the viewer. We witness illicit acts, yet are not informed as to how the participants *feel* about them. Thus, it is left to the viewer to make a judgment as to how *they* feel about the action of the piece. It is notable that the lust exhibited among the Israelites in *Dance* holds connotations with 'luxuria' or libido, which was considered one of the seven deadly sins by the church. However, in Renaissance times, many people began to regard 'volumptas', or sexual gratification, with the quality of lust (Hall 196). It is interesting that many Renaissance thinkers were actually coming to regard 'volumptas' as a virtuous and worthy pursuit. This is in line with Renaissance concepts of the glorification of the human, and the elevation of the physicality of man to a state of grandeur. This moral conflict between old church doctrine and modern Protestant thought was something Lucas could not ignore, and his decision to put the concept of the libido in the forefront to the biblical text indicates sympathy for modernized conceptions. Furthermore, the 'play' of the Israelites mentioned in Exodus 32:6 may refer to a sexual orgy of the Canaanites, who worshipped bull-calves as symbols of strength and fertility (Smith 296 Women). The stocky, bared body parts, especially knees, calves, and feet, found throughout *Dance* may be read as a visual reinforcement of this notion of strength. The pronounced breasts and belly of the female in the central panel further illustrates the notion of fertility. This depiction of the female was used to emphasize women's role as the mother of all humanity. Also notable is the repeated use of feathered caps in *Dance*. This fancy, decorative element may indicate frivolousness and absence of chastity (Smith 65 New Appraisal).

The aforementioned female figure in the central panel of *Dance* bears a striking resemblance to typical Madonna images of the time. Also, the figure is handing a piece of fruit to the male who is seated on the ground slightly below her, paralleling her form with the role of Eve in the fall of man. This depiction of woman as both 'janua diaboli' (the Devil's door) and 'mater et nutrix omnium' (spiritual mother and nurturer) is apparent in the centralized female form of Lucas's *Dance* (Smith 296 Women). This idea of woman as the simultaneous embodiment of creator and destroyer is a visual play on the Ave/Eva dichotomy which manifests itself throughout art history. Lucas was in tune with the popular conception of the sixteenth century woman as morally degenerate, yet physically miraculous. A woman was, on one hand, socially bound to uphold virginal ideals of purity, chastity, and virtue. However, the glorification of the female nude, as in Botticelli's *Venus*, highlighted the sheer erotic appeal of the disrobed female form. By depicting the centralized female figure as the embodiment of Ave/Eva, Lucas heightens the impact of his inverted composition and adds fuel to the fire of conversation his painting is intended to ignite.

The female in the centralized portion of the right panel of *Dance* holds further iconographical implications, largely due to the child beside her. With its references to Botticelli's *Venus*, the figure takes on connotations of eroticism and desirability. The child beside her attempts to hold onto a bird, which is about to escape in flight. The Christ child is sometimes shown holding a bird, (Hall 48), and it is possible Lucas is repeating the Virgin and Child implication of the center panel. Another possible interpretation of the bird about to fly away is the symbolic loss of the soul, which may result from the Israelite's revelry. It is notable that the child also wears a gold hoop earring, possibly indicative that he has remained pure by *not* participating in the construction of the golden calf. More likely, Lucas included this detail to give the boy a gypsy connotation. The combined imagery of the earring and the bird in relation to the boy could signify the child is tainted and wayward, in danger of losing his soul. Gypsy references are repeated in some of the headpieces and garments throughout *Dance*, creating an emphasis on the negative implications of gypsies as symbols of deceit and avarice, as depicted in Bosch's *Haywain*. Furthermore, the word 'gypsy' originates from Egypt, and the inclusion of such symbols in *Dance* may also allude to the Israelites' biblical escape from Egypt (Smith 308 Women).

The repetition of the form of a stick, sword, or staff provides additional erotic as well as alchemical references within the triptych. In the center panel, the child on the Ave/Eva figure's lap holds a stick, and his gaze lines up with the feather-capped gentleman in the right of the center panel. This man is gesturing back towards the center and holds what appears to be an empty sheath for a short sword, which holds obvious sexual connotation. The object may also be an empty tankard, which is a vessel also associated with female genitalia (Smith 300).



Figure 8. Lucas van Leyden, *The fall of Man (Adam and Eve)*, c. 1528 (drawing in black chalk). Hamburger Kunsthalle, Hamburg, Germany. (© Bildarchiv Preussischer Kulturbesitz / Art resource, New York)

Women). This image of hands grasping twigs, sticks, or swords finds its place in other Lucas works, most notably his 1530 drawing of *Adam and Eve* (Figure 8). The hand encircling the staff, sword, or stick in *Dance* thus becomes a metaphor for the phallus, and heightens the erotic connotations of *Dance*. The stick that bisects the left panel, the sheathed sword in the left of the center panel leading to the stick above it, and possibly the stick and sword held by Moses and Joshua in the background of the right panel may all be read with similar sexual significance. Weaponry, particularly swords, also holds an alchemical meaning. In alchemical imagery, weaponry signifies the action of the philosophical fire, especially at the initial stage of the opus (Abraham 214). Lucas may have used the repeating image of the sword throughout *Dance* as an alchemical metaphor for the new philosophies rising out of the Renaissance and Reformation. Once again, Lucas is commenting on the dawn of a new era in which people were awakening to newfound independent thought.

Perhaps the most iconographically potent imagery in the work, though, are the depictions of wine and wine vessels. During the age of the Reformation, one of the most heated

religious topics revolved around the spiritual value and effectiveness of the Sacraments (Silver 408). Lucas's inclusion of wine and wine vessels could well relate to the debate about the sacrament of communion occurring in the Church at this time. Communion was recognized at this time by Protestants and Catholics alike, but the debate over how it should be understood was a key point throughout the transition period of the Reformation.

Moreover, alchemically, wine was suggested by Aristotle to be symbolic of the union of male and female (Abraham 89). The view of wine as the vehicle to lust and fornication correlates with the interpretation of *Dance* as promoting eroticism. The appearance of the jug, wine drum, and act of pouring wine into a vessel by the male figure in the center panel create a repetition of motif. The shape of the vessels and drums in the middle ground area by the calf repeat the shape of the pedestal the calf rests upon, possibly correlating the wine as another form of idolatry. Importantly, the figure in the right portion of the center panel behind the Bosch inspired face holds a jug directly to his lips. Another figure in the left panel holds a cup right up to his mouth. One of the ideas that Martin Luther considered important was that the sacrament of communion should be considered more commemorative than miraculous, and for this reason, it was important for the wine cup to come into direct contact with the lips (Harbison 153). Lucas may have intended the motif of the wine vessel touching the lips to act as a visual commentary on the changes brought about by the Reformation in reference to the communion sacrament. Furthermore, the male figure in the center panel who pours out the sacrament into a cup could be interpreted as Aaron, the priest who constructed the idol for the Israelites. This may be a commentary on priests being vehicles to sin. A possible explanation of the male figure's action in the center panel may be explained by Exodus 32:20: "And he took the calf which they had made, and ground it to powder, and strawed it upon the water, and made the children of Israel drink of it". As seen by the biblical text, the Israelites actually consumed their idol after it was destroyed. Perhaps the figure pouring wine acts as a metaphorical foreshadowing of the consequence the Israelites will face when their idol is destroyed. The dark storm clouds which occupy the top quarter of *Dance* could also be indicative of the ensuing wrath of God.

Thus Lucas van Leyden's *Dance Around the Golden Calf* captures the tension and contradiction of the age in which it was produced. The use of inverted composition in *Dance* automatically subverses the biblical text in favor of more secular interests. One of the results of the tension of the Reformation age was to push more secular approaches to art. The color scheme of *Dance* juxtaposes vibrant hues of red against an otherwise earthy palate, setting up a subtle visual tension. The voluptuous, sensual forms observed show Italianate influence and add to the psychological implications of the piece. The expressionless faces reinforce Lucas's attempt with *Dance* to spark dialogue



about morality, as well as add to its intrinsic contradictions. Compositional decisions such as cutting off the bodies, but continuing the landscape reinforce this unification and division. The juxtaposition of the Ave/Eva as creator/destroyer observed in *Dance* creates an interesting dialogue about the conflicted view of women during this time. The erotic connotations observed throughout the iconography continue the secular vein of the work, showing the revolutionary philosophies which rose from the Reformation. Finally, the depiction of wine and wine vessels in relation to the Sacraments solidify this piece as a complex, multi-faceted dialogue on the changes in the church during the Age of Reformation.

### Bibliography:

- Abraham, Lyndy. *A Dictionary of Alchemical Imagery*. United Kingdom: Cambridge University Press, 1998.
- Arnason, Harry N. *History of Modern Art*. New York: Harry N. Abrams, Inc., 1998.
- Benesch, Otto. *The Art of the Renaissance in Northern Europe*. Great Britain: Geo Gibbons Ltd., 1965.
- Cuttler, Charles. *Northern Painting*. New York: Holt, Rinehart and Winston, Inc.
- Friedlander, Max. *From Van Eyck to Bruegel*. New York: Phaidon Publishers, Inc., 1956.
- Gilbert, Creighton. *History of Renaissance Art Throughout Europe*. New York: Harry N. Abrams, Inc., 1973.
- Hall, James. *Dictionary of Subjects and Symbols in Art*. Colorado: Westview Press, 1979.
- Janson, Anthony. *History of Art*. New York: Harry N. Abrams, Inc., 1995.
- King James. *The Holy Bible*. Utah: Intellectual Reserve, Inc., 2000.
- Leymarie, Jean. *Dutch Painting*. Switzerland: Skira, 1956.
- Mander, Carel. *Dutch and Flemish Painters*. New York: Arno Press, 1969.
- Silver, Lawrence A. "The Sin of Moses: Comments on the Early Reformation in a Late Painting by Lucas van Leyden" *Art Bulletin* Volume 1 No. 55 (March 1973): pp. 401-409.
- Smart, Alastair. *The Renaissance and Mannerism in Northern Europe and Spain*. London: Thames and Hudson, Ltd., 1972.
- Smith, Elise Lawton. *The Paintings of Lucas van Leyden-A New Appraisal, with Catalogue Raisonné*. London: University of Missouri Press, 1992.
- Smith, Elise Lawton. "Women and the Dance Around the Golden Calf" *Art History* Volume 15 No. 3. (Sept. 1992): pp. 296-316.
- Snyder, James. *Northern Renaissance Art*. New Jersey: Prentice Hall, Inc., 1985.

### Mentor comments:

Ms. Pease's mentor, Lynn Jacobs, points out the particular characteristics of the paper that make it an important piece of scholarship. She says,

Jennifer R. Pease's paper, "Lucas van Leyden's *Dance Around the Golden Calf: The Northern Triptych in the Age of the Reformation*", is a significant contribution to the study of art works in their historical context. In this paper, Ms. Pease argues that the stylistic and iconographic features of this triptych reflect the tensions of the age of the Reformation—that central moment during which Northern Europe shifted out of the middle ages into the early modern era. This was the time when Northern European artists first truly assimilated the artistic values of the Italian Renaissance, and the time when the values of Catholicism, and the universalistic Catholic church, came under serious attack. The stylistic and iconographic analysis in Ms. Pease's paper demonstrate how this triptych forms a site in which the various tensions and dualities of the Reformation era are negotiated.

Among the key points raised in this paper are the way in which the iconography functions as moral commentary through its association of the sins of the ancient Israelites with licentious behavior. The paper, however, is also sensitive to the multivalent character of the imagery and also considers how the painting uses the scene of the Golden Calf as an allusion to the Reformation debate both over the sacrament of communion and over the role of images in the church. Ms. Pease's paper also demonstrates cutting edge art historical methodology in its attention to the role of the triptych format itself as a vehicle of meaning and in its treatment of the triptych as a unified entity.

## BEURS IN THE HOOD: COMING OF AGE IN THE BANLIEUE

By Lauran Elam  
Department of Foreign Languages

Faculty Mentor: Dr. Kathleen Comfort  
Department of Foreign Languages

### Acknowledgements:

Thank you, thank you, thank you to Dr. Kathleen Comfort, the most helpful mentor ever. I could not have done it without you. And thank you to Dr. Hope Christiansen for all the help, advice, and offers to baby-sit over the years. Thank you to Dr. Kathleen Condray for all the advice, not just about German. Thank you to my family, in particular, my daughter, for being all the motivation anyone could need, and my grandparents, whose sacrifices have made my success possible. Nanny, we love and miss you.

### Abstract:

*In November of 2005, riots broke out throughout the housing projects located in the suburbs of the major cities of France. This focused worldwide attention on the largely Muslim immigrant communities in France, and on the failure of the French government to fully integrate individuals of foreign extraction, namely Beurs. The term "Beur" is the French word "Arabe" reversed by language called Verlan that plays on French in much the same way that Pig Latin plays on English. Today "Beur" refers to the children of North African immigrants living in France who are, for the most part, isolated to the housing projects of the banlieue (suburbs).*

*This article elaborates on what it is to be Beur through analysis of semi-autobiographical works by Beur authors, films focused on Beur characters, and newspaper and journal articles on the subject. In the media, Beurs have been portrayed as foreigners angry with France, and radical Islamists struggling against the West. However, these characterizations are not corroborated by primary sources. In reality, Beurs are trapped between two cultures, rejected by the French for their cultural background, and rejected by the first generation immigrants of their communities for abandoning traditions they have a hard time understanding, and acculturating to a more secular way of life within France.*

### Introduction:

I chose the topic of *Beur* literature and culture without really knowing what I was getting myself into. I suppose I was a little more knowledgeable than the average person on the street; I had at least heard the term "*Beur*" in my French Civilization course and had seen *La Haine* in the same class. I associated the term with the film and with the headscarf controversy that troubles the French educational system, but I went into this project without a coherent definition of what it means to be *Beur*. I had no idea how difficult and controversial such a definition would be.

The word *Beur* comes from *Verlan*, a language that plays on French in much the same way that Pig Latin plays on English. The syllables of words are reversed so that *Arabe* becomes *Beur*, but the word does not mean Arab. It is used to describe the French children of immigrants, particularly North African immigrants, but as I learned in the course of my research, the term does not apply to all individuals of this particular background, nor does it seem to apply solely to these individuals.

When I began this project its relevance was not as clear as it is today. Most people had no interest in the meaning of *Beur* prior to seeing the gangs of youths burning cars in the suburbs of France's major cities last November, and could have cared less about *H.L.M.'s*, *Habitation à loyer modéré*, rent controlled public housing, or their inhabitants. Suddenly amid all the media scrutiny, what it is to be *Beur* has become important. The youth of the *banlieue*, the French suburbs where most *H.L.M.'s* are located, were portrayed as everything from mere hooligans to Islamic extremists taking part in a worldwide clash with the West. The truth is far more complicated than that.

In gathering resources to formulate my definition of *Beur* literature and culture, I discovered that the majority of works by prominent *Beur* authors contain an auto- or at least semi-autobiographical element. These individual works are too

personal to paint an objective picture of what it is to be *Beur* on their own, but by finding common elements, similar circumstances, attitudes, and means of coping, one sees that the outline of a culture begins to emerge. I chose *Kiffe kiffe demain* by Faïza Guène because it is a current and extremely popular novel written by a fifteen year-old girl, herself a *Beur*, about a teenage *Beur* girl. *Béni ou le paradis privé* by Azouz Begag, one of the most prominent literary and cultural figures among first generation Frenchmen of Algerian extraction, was chosen because its plot centers around the experiences of a boy of a similar age and background as Doria, the narrator of Guène's book.

The films treated in this paper seem broader in their scope. The absence in the films of a first-person narrator, present in all the novels discussed here, paints a more generalized picture. *La Haine* seemed an obvious choice, since its plot mirrors the story of the riots, and also because it was the film that introduced me to the topic. *Camping à la ferme*, the story of juvenile delinquents from the *banlieue* sent to a country village to work as an alternative to incarceration, was chosen because the screenplay was written by Azouz Begag, and it provided a more up to date view of the issue, having been released just last year (*Béni ou le paradis privé* was published in the 1980's and is set in the 1970's). I chose *Samia* because I had seen it on *The Independent Film Channel* sometime before, not really understanding the significance of what I was watching. The film paints a clear portrait of a family of Algerian origin living just outside Marseilles. After watching the film again I discovered the novel, *Ils disent que je suis une beurette* by Soraya Nini, on which the film is based. The film focuses on the family, but the book addresses most everything from the family, to the neighborhood, the schools, and most every social institution through the highly critical eyes of its teenage narrator, Samia. Although Nini collaborated in the making of the film, the two differ greatly given that the book covers five years of Samia's life, and the film covers only one summer.

In order to put this all in context I looked at several journal articles and newspaper reports on everything from *Beur* literature and music to the state of Islam in France. The most valuable resource for this task proved to be *Francoscopie*, a book of demographic figures and commentary about sociological issues in France today. *Mémoires d'immigrés*, a history of North African immigration into France by Yamina Benguigui, also offered some insight. Much of what is said in the creative works is supported by statistics from the nonfiction. None of the sources written in the original French have been translated to English. The translations provided here are my own.

I questioned my task repeatedly, wondering how well a middle-class college student from Arkansas could really understand the situation of the children of immigrants living in housing projects a continent away. But after discovering the multitude of different ideas on the subject even within France, I decided that my distance from the situation might even work to

my advantage. In this paper I address the interactions of *Beurs* with various religious, cultural, social, and political institutions in order to ultimately define what it is to be *Beur* in the context of today's France.

### *Beur* Youth and Religion:

The French take great pride in their successful efforts at secularization since church and state were officially separated in 1905. But a pragmatic attitude toward religion in France can be traced even further back to Henry IV's "Paris is worth a Mass" and events like the Babylonian Captivity, when the Pope was moved to Avignon. Today the French hold to tradition with 75% of the population declaring that they have a religion, but only 50% declaring themselves believers (Mermet 271).

Catholics represent the religious majority in France today: 62%, with Muslims coming in a distant second at 6%. Protestants account for 2% of the population declaring an affiliation, and Jews account for 1%. The broad category of "other religions" makes up 2%, and the other 26% are said to be without religion (Mermet 271). These numbers, though, are somewhat deceptive. Of France's declared Catholics, 47% attend Mass only on holidays and special occasions, 24% attend from time to time, 7% one or two times per month, and 10% never go. The trend is even more pronounced among Jews; only 15% practice regularly, and only 4% practice occasionally (Mermet 273).

Islam suffers from the same lack of enthusiasm among its adherents in France. Estimates place the number of individuals of Islamic origin living in France at around 4 million. Of those 4 million, only 1.5 million declare themselves to be observant believers and only 20% report that they worship regularly ("Islam in France"). Islam's difficulties in France may be a result of the differences between the clergy and their prospective congregations. Of the 900 imams in France, more than 90% are foreign, educated on the Koran at universities in North Africa and the Middle East. The interpretation of the Koran taught in those areas often differs from the attitudes of French Muslim communities (Mermet 272). According to the article by Milton Viorst, the debate centers around the tendency of foreign-educated imams to advocate a religious government, and the tendency of those born or educated in France to reject such a notion in favor of secularism.

Youths of Muslim origin living in the *banlieue* seem to have a very cynical attitude about Islam and religion in general. Religion becomes the ill-explained and misused tool of authority figures. The young recognize that when it comes to piety, what the neighbors may think counts for more than genuine faith. And the double standard in place for girls and boys does not lend credibility to religion's cause. This phenomenon is well illustrated in several works, and directly articulated in *Ils disent que je suis une beurette* by Soraya Nini. In discussing what it is to be a woman in her family, and arriving at why her sister was disowned

immediately for the habitual offense of her brothers, going out unaccompanied with members of the opposite sex, Samia, the teenage narrator of the story offers this:

Il paraît que c'est la religion qui veut ça, et que chez nous la femme n'a pas le droit de faire telle ou telle chose, en bref, de vivre normalement! "Ça ne se fait pas chez nous," c'est la phrase magique pour dire qu'il lui faut absolument rester enfermée! De tout façon, il n'y a rien qui se fait "chez nous"! [Nini 123] (It seems that it's religion that wants this and that at our house a woman doesn't have the right to do this or that, in short, to live normally! "That isn't done at our house," are the magic words to say that she absolutely ought to stay shut away! At any rate, there's nothing that's done "at our house"!).

She obviously hates the lifestyle, but the thing that really offends her is the lack of explanation behind it.

C'est trop facile! On ne nous a jamais parlé de la religion. C'est vrai que mon father et ma mother prient, mais jamais on n'a su ce que cela voulait dire. Aucune explication, aucune histoire racontée sur la religion. Et d'un coup, on nous dit que c'est la religion qui nous interdit de vivre comme on le voudrait!... Pour moi, ce n'est pas ça la religion. Je ne sais pas vraiment ce que c'est... [Nini 123] (It's too easy! They never talked to us about religion. It's true that my mom and dad pray, but we never knew what that meant. No explanation, no story told about religion. And all at once they tell us that it's religion that forbids us to live the way we'd like! ...For me this is not religion. I don't really know what it is...).

It seems that, in the *banlieue*, religion is more about the forbidden, avoiding sin, than it is about belief, or avoiding hypocrisy. In Faïza Guène's *Kiffe Kiffe Demain* the young narrator tells us of an occasion during which her father responded violently when she displayed inappropriate interests and then justified his actions with religion. She recalls her friend giving her a poster of a boy band heartthrob and the reaction from her father when he found it hanging in her room.

Le soir mon père est entré dans ma chambre. Il s'est mis dans tous ces états et a commencé à arracher le poster en criant: "Je veux pas de ça chez moi, y a le chétane dedans, c'est Satan!" C'est pas comme ça que je l'imaginai le diable mais bon... Sur le mur vide il restait juste un tout petit morceau de poster avec le tétou gauche de Filip. [Guène 28] (That night my father came into the room. He worked himself up into one of his states and began to tear down the poster crying: "I don't want any of this at my house, there's the devil in it, it's Satan!" I didn't imagine the devil like that but okay... On the empty wall only a very small piece of poster remained with the left nipple of Filip.).

A lack of direct religious instruction in the families depicted in the stories discussed here contributes to an objectification of

religious belief among the young narrators. In the film version of Nini's book, *Samia*, the parents are depicted as praying often, though always alone in the bedroom, while the children watch television in the living room of the apartment. The only attempt at an explanation of religion is directed from the visiting Algerian aunt to the youngest son and daughter of the family. As she is preparing food in the kitchen she speaks of sin, and tells them that the path to god is narrow, like the blade of the knife she is using. The adolescent boy mocks and questions her immediately, at which point she tells him he is going to hell and the conversation ends. Later in the film, the title character, Samia, attempts to leave early from a wedding the family is attending only to be dragged back by her older brother. A veiled woman then knocks at the door of the family apartment and tells Samia that she has come, at the request of her mother. Her mother fears that Samia's rebellion will result in the sort of dishonor her older sister brought to the family by running away to live with her French boyfriend, and has asked this woman to come talk to Samia about religion. Samia slams the door in her face, and screams at her mother that she has no right. The cultural difference between generations in the *banlieue* makes it difficult for parents to explain Islam in the context of modern French society, particularly to young girls who view it only as a tool for repression and subjugation and resent the different standards set for each sex in the name of god. The works in question suggest that when religion is not explained, and is instead used as a threat, it loses credibility in the eyes of the young.

Islam can also lose some of its luster or appeal for children when compared to Western Christian spectacles like Christmas. As Béni, the narrator of Azouz Begag's *Béni ou le Paradis privé*, explains in the first paragraph of the novel: "Noël et son père barbu ne sont jamais rentrés chez nous." [7] (Christmas and its bearded father never visited our house.). He goes on to list the joys of commercialized Christmas including trees with lights, and chocolate lambs with the baby Jesus, explaining that they got nothing at their house:

Et tout ça parce que notre chef à nous c'est Mohamed. Dans son bouquin, il n'avait pas prévu le coup du sapin et des cadeaux de 25 décembre. Un oubli comme celui-là ne se pardonne facilement. On aurait presque envie de changer de chef, du coup, pour faute professionnelle! [Begag 7] (And all this because our own boss is Mohammed. In his book, he didn't anticipate the trick of the Christmas tree and presents on December 25. An oversight like that one isn't easily forgiven. We'd almost want to change bosses, as a result, for professional misconduct.).

Such feelings create more difficulties in Béni's 1970's *H.L.M.* where his father tries more vigorously to retain a sense of Muslim identity. But even Béni gets mixed messages in regard to religion. His father, extremely proud that Béni wins a gift for his knowledge of Christmas Carols at the company Christmas party in the beginning of the story, becomes enraged and beats

his son on the head with his shoe after he cries and defiantly sings a carol in order to convince the family of the need to buy their own Christmas tree. Even when some families take pride in the ability of their children to adapt to their new society, any genuine change, particularly within the home, elicits a violent reaction. As Yacine, the eldest brother in *Samia*, tells his defiant younger sister, Amel, "À la maison, c'est toujours le bled." (In our house, it's still the old country.). This is a popular sentiment among the first-generation characters of these stories, but one to which the protagonists often have a difficult time resigning themselves.

The term *Beur* may be a play on *Arabe*, but Arabs are not the only minority confined to the *banlieue*, and Islam is not the only religion of the suburban French ghettos. With the massive influx of North-African immigrants into France came a massive influx of immigrants from the Sephardic Jewish communities. Economic circumstances forced Jews and Muslims into the same *H.L.M.*'s, and cultural differences mean that the French view both as a threat to secularist identity. Just as cases have arisen challenging the rights of Muslim girls to wear veils in public schools, so too have there been conflicts regarding the rights of Jewish students to miss class on the Sabbath (Shurkin 161). But with only 15% of all those declaring themselves of Jewish origin regularly observing their religion (Mermet 273), Jewish and Muslim youths growing up in the *banlieue* seem to have far more in common than one might suspect. As reported by Glen Frankel in an article called "For Jews in France, a 'Kind of Intifada,'" acts of anti-Semitism in France, which have risen in recent years, are divided into professional acts of terrorism, like the firebombing of synagogues, and those perpetrated at random by gangs of street youths. France's Jews may have to worry about the attitudes of their Arab neighbors, but as the scenes from *Samia* and *La Haine* vividly illustrate, both groups face the same prejudices from the extreme French right and skinheads, and thus have some common ground.

In both *Camping à la Ferme* and *La Haine*, Jewish youths from the *banlieue* are depicted as delinquent, just as their Arab friends are. The boys in *Camping à la Ferme* are totally unaware that one of them is Jewish until they are all asked to work at a Catholic church as part of their community service. Religion does not have much sway on the boys. The only devout Muslim in the group is a Sicilian convert whom no one takes seriously, and apart from their refusal to work for the Christian church, one would not know that religion was even an issue. The response of the priest to their refusal, that they are all "children of Abraham" and so it should not be a problem, loses credibility vis-à-vis the townspeople, who do not react in such a tolerant manner. Religion really only seems to be an issue when a line is crossed, as when the boy's supervisor tries to find *Merguez* (sausage) or some other pork-free meat at the small town butcher, or when Doria's mother's boss denies her a break to eat after dark during Ramadan, and she has to hide dates in her pocket in order to continue to work.

To Samia in *Ils disent que je suis une beurette*, religion becomes convenient when she needs a reason for her absence from a friend's party. Too ashamed to tell her friend that she faces a beating at the hands of her brother, Samia excuses herself by saying that she has to stay home because it is a Muslim holiday like Christmas. Béni also has no problem setting religious significance aside when he presents the object of his affection, a beautiful French classmate, with the gold-pendant inscribed with a verse from the Koran that his mother has given him. She questions the significance of the gift, and we discover that for Béni, the fact that it was gold and valuable factored more in his decision than did the meaning of the verse.

Those who would characterize second-generation immigrants as differing from the rest of French society on the question of religion may be correct in that they do come from a different religious background, but by many accounts they also exhibit an innately French ability to set religion aside in the pursuit of secular goals.

### *Beur* Youth and Surrounding Cultures:

In many cases, *Beur* youths demonstrate an even more skeptical attitude toward the customs and traditions of their parents' countries of origin than they do toward religion. Ideas and practices contrary to equality between the sexes do not sit particularly well with children brought up in the relative *liberté* and *égalité* of the secular French educational system.

The youth of the *banlieue* are for the most part not cultural relativists. As evidenced in the novels and films, the cultural practices and mysticism of their North African parents, many of whom came from rural living conditions where religion or superstition factor into every aspect of life, often come across as exotic and, in the worst cases, as barbaric to their children. Most inhabitants from a Muslim background now living in France reside in the major metropolitan areas, about 90% (Mermet 272).

Their parents' country of origin is definitely depicted as a foreign country to the young female narrators of both *Ils disent que je suis une beurette*, by Soraya Nini, and *Kiffe kiffe demain*, by Faïza Guène. To Samia, Nini's narrator, her mother's Algerian sisters who come to visit the family appear so strange with their facial tattoos and braids that she nicknames them Geronimo and Cochise, because they remind her of Native Americans. She has long been aware of her mother's own less pronounced tattoos, but says of her perceived difference "Je sais pas, c'est pas pareil" [Nini 27] (I don't know, it's not the same.). Samia's mother brings this tradition of physical alteration, on a smaller scale of course, home to Samia and her younger sister in the form of a woman whom Samia calls a *sorcière*, or witch, and who cuts the cheeks of each girl with a razor blade while reciting a prayer and sopping up the blood with a raisin, which the girls are then forced to eat. The experience is extremely traumatic for both girls, who have to be forcibly held down to endure the ritual designed to protect their virginity for marriage. Her mother feels

it necessary after the family's eldest daughter, Amel, tired of beatings so severe from her eldest brother that she, the only gainfully employed member of the family, is forced to miss work, runs away to marry her French boyfriend. At this point, Samia's curiosity about cultural practices like these turns to resentment, and although the preceding episode is omitted from the film version of *Samia*, this resentment manifests itself in the movie when Samia and her sisters are sent to Algeria as punishment after Samia refuses to submit to a gynecological exam in the presence of her mother to prove her virginity.

Doria, Guène's narrator, also fears her parents' homeland, Morocco. During a trip there as an adolescent, a local approaches her mother saying,

Tu sais, Yasmina, ta fille devient une femme. Il faudrait que tu penses à lui trouver un garçon de bonne famille. Tu connais Rachid? [Guène 22] (You know, Yasmina, your daughter is becoming a woman. You ought to start thinking about finding her a boy from a good family. Do you know Rachid?).

Doria is not amused. She has met Rachid and adds that everyone calls him

"Rachid l'âne bête." [Guène 22] (Rachid, the beaten ass.). "She is not impressed by the fact that they have proposed to her an illiterate boy who, she tells us, lacks four teeth. To Doria, Morocco is also the country to which her father returned after divorcing her mother to marry a younger woman. He did so on the grounds that Doria's mother had not given him a son. Doria not only wishes to avoid the country on the basis of her own experience, but also because she resents the fact that her mother will not return out of shame for what has happened with her marriage.

It is easy to see why Algeria or Morocco would be sources of resentment for acculturated French girls of North African heritage when Samia's eldest brother, Yacine, tells his sisters in the film, *Samia*, that at their home they are "toujours au bled." (still in the old country). The implications of that statement on the girls' lives at home are grave. In the family's tiny apartment, the girls are required to do all the cooking, cleaning, and household chores. They are even required, to Samia's great annoyance, to clean the rooms of their brothers, none of whom are employed. They risk beatings, always at the hands of Yacine, eager to fill the role of strict Algerian father during the ailing family patriarch's convalescence with tuberculosis, for acts of insubordination such as refusing to serve him at the dinner table after a beating.

The women of Samia's family are not the only ones to live in fear of fathers and older brothers. In *Béni ou le Paradis privé*, Béni's mother also fears corporal punishment after she loses the family's grocery money while shopping with Béni in the market. She becomes hysterical, telling her young son she fears for her life, but manages to escape physical consequences by feigning a depressive illness until she can come up with the money. Although

in this story, the physical aggressions are not directed so exclusively toward the women of the household. Here everyone, with the possible exception of Béni's father, has someone to fear.

In *Kiffe kiffe demain* Doria's upstairs neighbor, whom she hears being beaten regularly by her father and older brother and whom she witnesses being dragged into the apartment by her hair when she comes home late from school, finally disappears from the neighborhood entirely. Later, rumors that she is living downtown with a French boyfriend begin to circulate. Doria is supportive, but much of the talk from the older generation of the neighborhood is disapproving. The people do not directly condone the beatings, but seem to look the other way.

However, *Beur* youth are not alienated from the culture of their parents' home countries only to fall into the welcoming embrace of a hospitable French culture. Religious and class issues guarantee that those who have grown up isolated in immigrant communities are always distinguished as separate. The French Government may strive for secularism, but in a country where Catholics still constitute 62% of the population, someone from a non-Christian background is sure to feel left out. Taking into account that 26% of France's population declares that they have no religion, the number of individuals who are "Christian" by cultural tradition may be significantly higher (Mermet 271). Many atheists may not believe in the meaning of Christmas, but they still trim a tree and buy gifts for their children even if they do not go to Mass.

The distance between the culture of the *banlieue* and that of France as a whole is nowhere more obvious than in *Béni ou le Paradis privé*. Béni fights for a Christmas tree and carols in the apartment only after having attended the Christmas party of his father's company. There is a portion of the festivities during which the event's MC presents the workers' children with gifts. He announces that the band will play a Christmas carol, without words, and that the first child to come to the stage and name the song will receive a gift. Béni sits quietly with his family, never having heard the songs being played until a mother seated in front of him whispers the name of one of the songs to her son who is too shy to approach the stage. Béni seizes the opportunity and heads toward the stage to the amazement of both the crowd and his family. Once there, he announces into the microphone that the title is "Le Sang des cerises" (The Blood of Cherries). When the MC tells him his response is incorrect the crowd pleads his cause and the MC gives him the prize although the actual title is "Le Temps des cerises" (The Time of Cherries) [Begag 16]. At this time, Béni's father is proud, but his son's continued interest in the holiday later makes him uncomfortable.

The culture of religion is not the only aspect to separate *Beurs* from the rest of society. Just as most Americans do not know precisely what to do when they are confronted with a *bidet* for the first time, neither does Béni. He is staying in a hotel room with his older brother and cousin following the death of a family member, and the cousin must explain to him why his excrement

will not flush. When the cousin, perhaps more acculturated than Béni or his brother, Nordine, gets angry and tells Béni he is crazy, Nordine quickly denies having told Béni that the *bidet* was the place to go, disguising his own ignorance.

The French are often heard to complain of the lack of willingness to acculturate on the part of immigrant families, but separated by religious and financial circumstances, and isolated into communities predominately inhabited by other immigrants, opportunities to acculturate outside of school and work are few and far between.

### **Beurs and Authority: French and Otherwise:**

France welcomed the wave of foreign workers who came to ease the country's labor shortages during the post WWII economic boom, but France was ill prepared for permanent immigration. The consensus seemed to be that these workers would arrive, earn enough money to ameliorate their conditions at home, and depart within a few years, returning forever to the place from which they had come. Efforts to acculturate these individuals seemed unnecessary in light of the transient nature of their circumstances. Whether or not they should have, no one anticipated the flood of immigration that came when immigrants brought their families to experience the relative prosperity to which they had grown accustomed ("H.L.M.").

France was even less prepared for the birth of what was termed a second generation of immigrants, particularly after the economic troubles the country suffered beginning in the 1970's. The mere fact that they were called second-generation immigrants illustrates the lack of understanding of their condition on the part of both the French government and the general population. These are children born in France, who speak French, attend French schools, watch French television, buy French products, and by and large want recognition that they are citizens of their country of birth. They may be the children of immigration, but they are by no means immigrants themselves (Benguigui 131).

Perhaps one of the most significant indications of France's lack of preparation for or understanding of its immigrant situation is the creation and maintenance of the 'H.L.M. (*habitation à loyer modéré*)' system. The concept of H.L.M.'s was born in 1950 to solve the housing shortages created not only by immigration, but also by the baby-boom and a post-war rural exodus that has shaped France's economy and demographic figures ever since. These structures were designed as temporary housing, constructed very quickly in a uniform, perfectly rectangular fashion so that the cranes they used could run along straight tracks, requiring less time and effort for construction. These high rises grew prolifically dotting the landscape of the *banlieue* (suburbs) of most major French cities ("H.L.M").

Conditions in the H.L.M.'s have improved somewhat since the implementation of the system, in that the average number of rooms in each apartment has increased, the average number of

inhabitants per dwelling has decreased, and amenities such as private toilets are now standard. But the French are still suffering from the consequences of their faulty planning ("H.L.M."). Alan Riding points out in an article for the *New York Times* that these communities were often so hastily constructed that they lack the proper infrastructure like wide, well lit streets to allow for easy policing. Problems like these become tragically apparent when the *banlieue* erupts into chaos as it did last autumn. That situation garnered national attention when it spread from Paris into other cities in France, and even beyond into other parts of Europe, but it was merely an escalation of regular occurrences in these isolated communities where crime rates are high ("Banlieue").

As stated above, this public housing was designed to be temporary, but as of 2003 the average H.L.M. family had lived in the same apartment for eight years (Mermet 177). Urban poverty rates, two times higher than the national average, widespread unemployment, and a relative lack of upward social mobility make it very difficult for families dependent on government aid to improve their situations and to move to private residences (Mermet 177, 220).

This isolated environment has bred delinquency. Unemployment is highest among non high-school graduates and young people ages 15-29 (Mermet 294-5), two groups that encompass many H.L.M. inhabitants. Here again the government was ill prepared. The problem of juvenile delinquency among the children of immigrants and those living in the *banlieue* did not become apparent until the 1970's. In Yamina Benguigui's *Mémoires des immigrés* one official was quoted on the subject of Beur youth as saying:

Je sentais qu'on avait créé une bombe à retardement, c'est à dire qu'on ne se rendait pas compte que les problèmes étaient posés: les enfants étaient très mal scolarisés et on fabriquait donc des délinquents. [Benguigui 131] (I feel that we created a time bomb, that is to say that we did not count on the problems arising: the children were very poorly educated and thus we made delinquents.)

Unemployment, educational difficulties, and delinquency are themes that are rampant throughout novels and films on the subject of the *banlieue*. All three are evident in *Samia*, and *Ils disent que je suis une beurette*. The film opens with a fourteen or fifteen year-old Samia being asked by a teacher to make a list of three careers she might like to try, since her future must be decided at that point. Samia has a rather hopeless attitude about the whole situation, saying that she really would prefer to do nothing. The situation is far more elaborate in the book, with Samia hopping from school to school, never happy with the attitudes of the teachers toward her, and pained by comparisons drawn between her and her sisters, all far more successful students. As the least academically successful sibling of the family, Samia is drafted by her mother into helping with the housework. Her mother's attitude in the book is thus rather



contradictory. French students are only legally required to attend school until the age of sixteen, but when Samia leaves one school under difficult circumstances her mother demands that Samia not drop out, that she get at least some sort of technical diploma. Then after searching for a new school for several days, she makes Samia late for her first meeting with the director of the school by keeping her home to do housework.

French schools are controversial enough; their treatment of the issue of the right of Muslim girls to wear headscarves to class received almost as much attention as the recent riots. Their attitude about the headscarf is difficult to pin down, with the national ruling not officially banning it, but essentially never resolving the decision as to whether or not such overt religious symbols are "disruptive" to the administration of a school. The issue is so complicated as to merit its own book or paper, but is not even mentioned in any of the books or films treated in this paper. However, it does seem symptomatic of a lack of willingness to compromise culturally on the part of the French educational system, notorious for beginning each history book with "Nos ancêtres les Gaulois" [Ha 103] (Our ancestors the Gauls). This kind of environment does not breed parental involvement on the part of immigrant families, but it may account for lower mosque attendance in France than in countries like Germany that, as Peter Schneider reports, permit optional Islam classes in schools just as they do classes on Christianity.

The attitude of teachers toward their *Beur* students in France seems less than welcoming. Doria tells of her school being placed in a sort of lockdown after an attack on a teacher. Samia has a series of teachers who not only quiz her about her family circumstances in front of the entire class, but even go so far as to try to humiliate her for not paying attention by forcing her to come to the front of the class to tell about "her" country, Algeria, a place she really knows nothing about. After the last incident, Samia stops doing homework entirely in an effort to fail and to be transferred to another school.

In this atmosphere that hardly embraces multiculturalism, it is easy for Samia and her siblings to go unsupervised when it comes to their schoolwork and activities. Because Samia's mother is unable to speak French, Samia can give most any excuse why she does not have homework or need to be in school that day, and her mother cannot verify the truth. The same is true for the activities of her youngest brother, Foued. Of course, his actions are not subject to the same kind of scrutiny that the actions of his sisters endure, but things catch up with him in the novel when the police storm and search the home without warning, looking for and finding evidence that he is part of a group of boys who have robbed and trashed several stores in the neighborhood.

Productive activities appear hard to come by in the *banlieue*. Neither of Samia's older brothers is able to find work. Yacine leaves for Paris searching for a job, but to no avail. He is also seen

calling to inquire about jobs in the film, but again without luck. The hopelessness of the employment situation results in an attitude of tolerance for petty crime or dubious activities among the characters of the stories. Doria idolizes Hammoudi in *Kiffe kiffe demain*, despite the fact that she knows he is a drug dealer, and she is outraged more by the arrest of Zohra's son than by the fact that he too was dealing drugs. The tendency to look the other way manifests itself when Samia talks about Malik disappearing for days at a time: "Je ne sais pas ce qu'il fabrique vu qu'il ne travaille pas, mais on ne le voit pas souvent." [Nini 39] (I don't know what he does seeing that he doesn't work, but we don't see him that often.). Even Béni, the narrator who prides himself on his acculturation and stellar academic performance in a way that Samia and Doria would never do, falls in with a group of what one might term street thugs when his family moves to a new *H.L.M.*

Several institutions have been put in place to combat juvenile delinquency in the *banlieue*. As Paul A. Silverstein reports in his article "Islam, Soccer, and the French Nation State," "following the election of François Mitterrand to the office of president in 1981, immigrant associations were legalized and legitimized. These groups offered Islam as an alternative to drugs and crime in the *H.L.M.*'s. They provide services like finding jobs in mosques for the unemployed and organizing summer camps for urban youth, but given the fact that these groups often receive international funding, more specifically from the Middle East, they often incite fear of radical Islam and terrorism among the French.

Silverstein contends that this fear of a step away from secularization and toward foreign influence prompted the French government to create alternative summer camps and build sports facilities in the *banlieue* in an effort to win the hearts and minds of young Beurs. The French promoted urban, or rather suburban, revitalization programs backed by sports stars like the soccer player Zinedine Zidane. The son of Algerian immigrants, Zidane grew up in the *banlieue* outside Marseilles, and after making a name for himself served as a sort of poster-child for acculturation. But with the sports stars came the sponsors and advertising. It seemed that the government, and the international corporations, namely shoe makers like Nike and Adidas, wanted to propose sports as an alternative to religion. The most overt manifestation of this attitude came with a series of ads placed in the Paris Metro that played directly on Islam and its conflict with the French government. Two of the most controversial phrases displayed in the ads, always with photos of athletes hardly featuring the product being sold, were "*Il n'y a pas qu'un Dieu. Il y en a onze.*" (There is not but one god, there are eleven.). This slogan plays directly on the line from the Koran: "There is but one god, and Mohammed is his prophet." This suggests that on the soccer field the players are the gods. Following the success of the Parisian soccer club, P.S.G., "*Aucune loi ne vous empêche de porter un maillot P.S.G. à l'école,*" which refers to the headscarf

controversy, pointing out that "No law stops you from wearing a P.S.G. jersey to school."

These programs, both state and privately funded, have experienced varied rates of success. The importance of soccer to the youths in the stories is obvious; it is Béni's only means of making new friends, and Samia is depicted at soccer games throughout the film. But these programs have also prompted resentment and questions about why athletic instead of academic success should be so vigorously promoted in these areas. In *La Haine*, Hubert, the avid boxer, suffers the direct consequences of this resentment when the community gym he has helped establish is completely burned, most likely by his friends, in the course of a riot that seems all too familiar. Doria, too, resents her family's social worker, who proposes that she go to one of these summer camps. Not only is she cynical, she is afraid to leave her mother to cope on her own.

When these programs fail to curb crime in the *banlieue*, the French government takes far more drastic measures, calling in the armored riot police now so familiar from photos of strikes and the riots of last fall. After a series of terrorist attacks in the mid 1990's the government went one step further, adding regular security patrols through these neighborhoods to enforce already legal blanket identity checks for anyone who looks "Beur" (Silverstein 31).

The *Beurs* of these stories are obviously targeted by the police, legitimately or not, particularly the males. Béni suffers from racial profiling after being caught making mischief with his more delinquent friends. He assumes he can speak rationally to the police officer and learns the lesson that his appearance, that of a *Beur*, counts more for police than his words ever will. The boys of *La Haine* feud constantly with the police, and *Camping à la Ferme* illustrates the hypocrisy of the government programs designed to rehabilitate young delinquents from these neighborhoods. One striking example comes in *Samia* when Yacine is forcibly stopped from passing the spot where a young man of African origin is being violently arrested by a group of black-clad police officers.

Problems with the police are symptomatic of the contempt with which most government agencies and representatives treat immigrants and their children. Benguigui's *Mémoires des immigrés* says that young *Beurs* cannot tolerate their parents attitude: "Travaille et tais-toi." [Benguigui 134] (Shut up and work.). And while Béni, our 1970's era *Beur*, reacts somewhat passively to the policemen who mock him when he goes searching through the market for the money his mother has lost, Samia and Doria react with equal contempt to anyone who dares treat their parents disrespectfully. Doria mocks her social worker constantly, resenting even the woman who replaces the despised man who told her that he had never met "des gens comme nous avec un enfant seulement par famille." [Guène 18] (people like us with only one child in the family). Samia is by far the most assertive

in *Ils disent que je suis une beurette*. She resents anyone who disrespects or talks down to her mother because she cannot speak French. She clashes with everyone from her school principal to a nasty bureaucrat to the police officers who come to arrest Foued.

### Who *Beurs* Are, as Opposed to Who They Are Not:

The question of *Beur* identity really emerged in France in the late 1970's and early 1980's after a series of equal rights marches, and the establishment of *Beur* media, like *Beur TV* and *Beur FM*, caught the attention of the public. Things proceeded as the next couple of decades witnessed the emergence of *Beur* writers and public figures who made a particular impact with their autobiographical or semi-autobiographical works. But international attention, or at least Western international attention did not really focus on the issue until the terrorist attacks, and rising tensions in the Middle East, described in the article by Milton Viorst, brought *Beurs* to the forefront starting in the 1990's.

The riots of fall 2005 played out on television screens around the world, and the media were quick to view this as yet another example of Islamic radicals lashing out at the West. A few articles appeared trying to paint a more realistic picture of what was happening and the reasons behind it, including one article by Alan Riding appearing in the *New York Times* with the headline "Artists have been sounding the warning bells for years." Indeed, the incident that sparked the riots, involving two youths electrocuted while being pursued by the police, mirrored the scene, involving a young man hospitalized in a coma after being beaten by the police, that sparked the violence in *La Haine*, a film released nearly a decade before. But ask most anyone on the street in an average American city about the riots and the typical response will be something along the lines of "Oh, you mean the Muslims."

But it is most definitely not that simple. The idea that *Beurs* were behind the riots is not at all in dispute, but the question that remains is what exactly it means to be *Beur*. It is not merely *Arabe* backwards. The books and films discussed in this paper, coupled with the statistics on France's current demographic, make it obvious that being *Beur* is no longer simply a matter of being Muslim or even of being Arab. In France, 15.6% of families live in *H.L.M.*'s, the settings for the books, the films, and the scenes of the riots (Mermet 177). Admittedly, the concentration of immigrants and their children is higher in *H.L.M.*'s, but given the fact that only 6% of France's population is of Muslim extraction, the idea that the pervasive delinquency and violence of the *banlieue* is merely a Muslim problem bears no credence. Today being *Beur* is a matter of being young, having grown up in a particular set of sociological circumstances, being adapted to that set of circumstances, and ultimately it is a matter of lacking any other suitable label.

In years past *Beur* was applied as a blanket term for all French-born children of Maghrebi parents. The consensus seemed to be that being *Beur* was simply about being trapped between two cultures and belonging to neither. But as the issue of *Beur* identity was further explored, the variety of identities within the immigrant community began to be recognized. An article from *Middle East Report* by Joan Gross, David McMurray, and Ted Swedenberg entitled "Rai, Rap, and Ramadan Nights: Franco-Maghrebi Cultural Identities" did a very good job of chronicling the differences among groups within *Beur* culture from those who fully embrace being French to those who still think of themselves as essentially foreign. The same article asserts that the term "*Beur*" is not acceptable to the Franco-Maghrebi population, and indeed "*Beur*" in general is offensive to some; in certain condescending contexts, it is offensive to most.

But for many, particularly now that rap music has emerged as the primary means of musical expression, "*Beur*" is something of a badge of honor. In much the same way that the impact of the "N-word" has been partially weakened by African American artists who have used it often in their music, changing its meaning from insult to familiarity, "*Beur*" has become an identifier for the young, disillusioned, anti-authoritarian youth of the H.L.M.'s. In *Ils disent que je suis une beurette* Foued raps about being *Beur*. *Beur TV* and *Beur FM* embrace the word. It seems to be the only adjective that describes the culture of their audience, no longer simply a mixture of French and Maghrebi, but now strongly influenced by foreign, particularly American cultures.

When asked to speak on the *Beur* literary movement, Lotfi Bennour, an English professor at the Université de Belfort in France who immigrated from Tunisia, answered that *Beur* is only used by the youth of the second and third generation born in France, by rappers, or by politicians and those who wish to encourage voter registration; thus *Beur* is not a term can be applied to someone like Azouz Begag, author of *Béni ou le paradis privé*. Begag, a part of the second generation who went on to become an important author, and who today serves as the *Ministre délégué à la promotion d'égalité des chances* for the French government, is someone that a self-proclaimed *Beur* might term a "*Beur de service*," that is to say someone who is integrated and successful within the French system and thus does not suffer the same consequences as others who share the same cultural background ("*Beur*"). To Bennour, *Beur* literature does not exist. It must either be grouped with French literature, given that the authors are French, or as post-colonial literature, given that it is a product of the post-colonial conflict.

But *Beur* culture is not simply the product of immigration or half-hearted integration. The attitudes reflected in the films and novels treated in this paper are part of a larger trend among the youth of France toward what *Francoscopie* terms counter culture with its own coded languages and influenced more by the media than by family or other institutions like schools (Mermet

99-101). When asked about his national identity, Zinedine Zidane always preferred to respond that he came from Marseilles or from the particular neighborhood where he grew up (Silverstein 41). Much like Zidane, the young men portrayed in *Camping à la ferme* identify with the neighborhoods they have come from, offering their neighborhood along with their parents' country of origin when asked where they were from by the film's token French bigot. One of the funniest and most poignant moments of the film comes when the last young man replies simply "France." When the bigot asks incredulously how that could be, the reply he gets is "*Par la chatte de ma mère*." This phrase translates as "through my mother's pussy."

In *Ils disent que je suis une beurette* Samia and her sisters create a coded language, even more specialized and complex than Verlan, tailored precisely to their household, mixing English, French, Verlan, Argot, and Arabic in order to speak freely and operate under the limited radar of their parents and older brother. Verlan has become rather mainstreamed over the past few decades, and so in order to stay on point and still be useful in an exclusionary way, words are even being re-invented so that for some "*beur*" has become "*reub*" or "*reubeu*" ("*Beur*"). It is clear that this individualized expression is an adaptation to living between two cultures: avoiding the scrutiny of the French teachers, police, and social workers while avoiding the scrutiny of foreign parents and thus avoiding authority in its entirety.

Lotfi Bennour says that some youths object to the idea that sports and music, more specifically rap, are offered as the only means of success for those who have grown up in the *banlieue*. And this may be true, but these stories make it obvious that these kids love their soccer and their rap. Béni is only able to infiltrate the group of friends he meets in his new neighborhood after he shows them a newspaper clipping featuring him winning an award for playing soccer. Samia dances wildly at soccer matches with her friends in the film, and in the book she makes up for the fact that she has to tell her teacher that no one in her family works by saying that her brother is an athlete. Doria ponders the street rappers she encounters at the festival in her neighborhood, concluding that even if the subject matter is violent and not necessarily intellectual, the medium at least does its part to promote equal rights by allowing neighborhood girls to sing back-up.

American culture is everywhere in literature and film concerning *Beur* subjects. The soundtracks may be laced with rap and *rai* but feature American music prominently. And Doria relates everything to American television, from *The X Files* to *Little House on the Prairie*. But French culture is everywhere too. These kids know French literature. They learned it in school. The history of the Revolution and other major events is just as ingrained in their psyches as it is in the psyche of anyone else educated in the French system. They have weak authority figures on both sides competing for their cultural allegiance and thus are free to pick and choose from a variety of cultures to which they

are exposed in defining their own cultural identity. If they had the means to escape the hopeless bubble of the *banlieue* and attain a French standard of success, then we might see a more voluntary acculturation, but most of them do not. By the same token, if they were ready to embrace Islam or the cultural practices of their parents wholeheartedly, we might see a more voluntarily segregation, an outright rejection of alternative cultures, but most of them are not.

*Beurs* are French by definition, whether they like the French or not, and whether or not the French like the idea. *Beur* literature exists as a part of French literature. Many want to lump it in with Francophone literature, or that literature written in French by and about people living in former colonies, but if it is written in France, by French citizens educated in France, about French citizens living and educated in France, it is logical to consider it French. The existence of *Beurs* may be a consequence of post-colonial policy, but they are at least a generation removed from living in those colonies. They are every bit a part of hexagonal France, even if they are in some ways still excluded. Riots and rage at circumstance are not symptoms of what happened to their ancestors in their countries of origin, they are symptoms of what happens to them daily in France. If the literary authorities cannot abide the inclusion of this genre of literature into the umbrella of French literature, then they need to recognize the genre as something entirely new, a change within France, a literature that in all objective respects is French, not Francophone, but is in some way different. That would mean that France would be officially multicultural, that dirty word that strikes fear in the hearts of many in France. France faces a tough choice about what to do with its citizens living in this situation, and unless they can act decisively to remedy the situation, they face manifestations of discontent larger than what they dealt with, or what they diffused but had not yet dealt with last November. In the last lines of *Kiffe kiffe demain* Doria imagines the revolt she will lead through the *banlieue* against the system that has oppressed her mother and sent her to beauty school against her wishes in the name of education. She recalls the Napoleon quote, "À tout peuple conquis, il faut une révolte." [Guène 188] (For all conquered people, there must be a revolt). But she adds:

Mais ce sera pas une révolte violente comme dans le film *La Haine* où ça se finit pas hyper bien. Ce sera une révolte intelligente, sans aucune violence, où on se soulèvera pour être reconnus tous. Y a pas que le rap et le foot dans la vie. Comme Rimbaud, on portera en nous «le sanglot des Infâmes, la clameur des Maudits.» [188-9] (But this will not be a violent revolt like in the film *La Haine* where it doesn't finish very well. It will be an intelligent revolt, without any violence, where we all stand up to be recognized. There's more to life than just rap and soccer. Like Rimbaud, we carry in us "the sobs of the infamous, the clamor of the confounded.")

She says, and sounds quite French saying, that she has "de forts élan républicains" [Guène 189] (strong republican urges).

### Works Cited:

- "Banlieue." *Wikipedia, the Free Encyclopedia*: March 30, 2006. *Internet*. April 6, 2006. <http://en.wikipedia.org/wiki/Banlieue>.
- Bennour, Lotfi. Email interview. March 1, 2006. "Beur." *Wikipedia, l'encyclopédie libre*: April 1, 2006. *Internet*. April 6, 2006. <http://fr.wikipedia.org/wiki/Beur>.
- Begag, Azouz. *Béni ou le paradis privé*. Paris: Éditions du Seuil, 1989.
- Benguigui, Yamina. *Mémoires d'immigrés: l'héritage Maghrébin*. Canal + Éditions, 1997.
- Camping à la ferme*. Screenplay by Azouz Begag, Jean-Pierre Sinapi. Dir. Jean-Pierre Sinapi. Perf. Roschdy Zem, Rafik Ben Mebarek, Julie Delarme. Pan Européenne Production, 2005.
- Frankel, Glenn. "For Jews in France, a 'Kind of Intifada'." *Washington Post Foreign Service*: July 16, 2003. *Internet*. March 26, 2006. <http://www.genocidewatch.org/AntiSemitismJuly16France.htm>.
- Gross, Joan. McMurray, David. Swedenburg, Ted. "Rai, Rap, and Ramadan Nights: Franco-Maghribi Cultural Identities." *Middle East Report* 178 (1992): 11-16+24.
- Guène, Faïza. *Kiffe kiffe demain*. Paris: Hachette Littératures, 2004.
- Ha, Marie-Paule. "From 'Nos Ancêtres, les Gaulois' to 'Leur Culture Ancestrale': Symbolic Violence and the Politics of Colonial Schooling in Indochina." *French Colonial History* 3 (2003): 101-17.
- "HLM." *Wikipedia, the Free Encyclopedia*. January 1, 2006. *Internet*. April 6, 2006. <http://en.wikipedia.org/wiki/HLM>.
- "Islam in France." *Wikipedia, the Free Encyclopedia*: March 21, 2006. *Internet*. March 26, 2006. [http://en.wikipedia.org/wiki/Islam\\_in\\_France](http://en.wikipedia.org/wiki/Islam_in_France).
- La Haine*. Screenplay by Mathieu Kassovitz. Dir. Mathieu Kassovitz. Perf. Vincent Cassel, Hubert Koundé, Saïd Taghmaoui. Canal +, 1995.
- Mermet, Gérard. *Francoscopie: pour comprendre les Français* 2005. Paris: Larousse, 2004.
- Nini, Soraya. *Ils disent que je suis une beurette*. Paris: Fixot, 1993.
- Riding, Alan. "In France, Artists Have Sounded the Warning Bells for Years." *The New York Times*, November 24, 2005. *Internet*. March 24, 2006. [www.nytimes.com](http://www.nytimes.com).
- Samia*. Screenplay by Philippe Faucon, Soraya Nini. Dir. Philippe Faucon. Perf. Linda Benahouda, Mohamed Chaouch, Nadia El Koutei. Canal +, 2000.
- Schneider, Peter. "The New Berlin Wall." *The New York Times*: December 4, 2005. *Internet*. April 14, 2006. <http://www.nytimes.com>.
- Shurkin, Michael Robert. "Decolonization and the Renewal of French Judaism: Reflections on the Contemporary French Jewish Scene." *Jewish Social Studies* 6 (2000): 156-76.
- Silverstein, Paul A. "Sporting Faith: Islam, Soccer, and the French Nation-State." *Social Text* 65 (2000): 25-53.
- Viorst, Milton. "The Muslims of France." *Foreign Affairs* 75 (1996): 78-96.

### Works Consulted:

- Lay-Chenchabi, Kathryn. "Writing for their Lives: Three Beur Writers Discover Themselves." *Mots Pluriels* 17 (2001). *Internet*. March 24, 2006. <http://www.arts.uwa.edu.au/MotsPluriels/MP1701kjc.html>.
- Simon, Patrick. "L'invention de l'authenticité: Belleville, quartier juif tunisien." *Revue Européenne des Migrations Internationales* 16 (2005): 9-41.
- Valdman, Albert. "La Langue des faubourgs et des Banlieues." *The French Review* 73 (2000): 1179-1192.
- Vassberg, Liliane M. "Immigration maghrébine en France: l'intégration des femmes." *The French Review* 73 (1997): 710-20.

**Mentor's Comments:**

Dr. Kathleen Comfort made the following remarks about Ms. Elam's work:

I have known Lauran Elam for three years now, when she was in major-level courses in French that I have taught. Her work has always been outstanding and her grades were consistently the highest in each class. In spring 2005, she took the Business French course for honors credit, for which she wrote a twenty five page research paper, in French, on Wal-Mart's market place in French-Speaking Canada. Her work was truly exceptional, and the amount of research she did was way beyond what I am used to seeing from undergraduates. When she approached me about being a director of her honors thesis, I was delighted because I knew that not only would she do an outstanding job, but that I would benefit from the experience as well. As it happens, I was correct on both counts; Lauran began researching her project last fall, well ahead of her peers, and worked tirelessly to find the ideal presentation of her ideas. Rather than seeking out sources to "fit" her ideas, she found herself calling into question her understanding of France and its people. To my mind, that is the hallmark of the scholar.

What sets Lauran apart from the vast majority of her peers is her work ethic and her desire to produce the best work she possibly can. Her project was by no means easy; the novels and the films she examined contain a large amount of French slang, much of it not found in standard dictionaries. In fact, most of the novels have not been translated into English. What is more, one of the movies, a new release ordered from France, did not contain English subtitles. She did her own translations of the quotations from the primary works, an undertaking that would challenge those with much more experience translating from French into English. Again, her work is on par with professionals in the field.

The originality of Lauran's project lies in her decision

to analyze both novels and films; to date, no scholar in the field of contemporary French literature has brought together these particular novels and films. Lauran's examination of the coming of age novel in *Beur* is, in many respects, groundbreaking. Working from a variety of secondary sources (internet sites such as *Beur FM*, blogs, sociological and anthropological studies, as well as literary criticism articles), she tackled the thorny question of what it means to be *Beur*. I am certain that any one of her chapters could stand alone as articles worthy of publication in professional journals in French literature. It is also significant that Lauran has stated that she hopes expand on her thesis in the future; I can easily see her continuing the project, with other novels and films, as part of a doctoral dissertation.

The topic Lauran chose is a timely one for those who study French language and culture. As was widely reported in the media, France was shaken by several weeks of rioting last fall in the *banlieues*, the equivalent of our inner cities. The cause of the rioting is complex, but at its core is the marginal status of young French citizens of North African extraction. French by birth, they are met with discrimination at every turn based on their ethnicity. The events were a wake up call for the government of Chirac and for French society at large, and opened up a debate on the sensitive topic of race and economic status, topics that have been nearly taboo in a society adheres to the slogan of the Revolution of 1789, *Liberté, Égalité, Fraternité* (Liberty, Equality, Fraternity).

Lauran's thesis invites to draw comparisons to the changing face of Americans, a growing percentage of whom are children of immigrants from Central and South America. Almost every day, there are newspaper articles on the efforts being made to better meet the needs of those from culturally diverse populations. Like the French, Americans must confront the assumptions and stereotypes about culture and ethnicity that are unfortunately all too prevalent.

## THE SOUNDS OF JAPANESE NOISE: FIRST GENERATION OF JAPANESE NOISE-ARTISTS

By Ana María Alarcón Jiménez  
Department of Music

Faculty Mentor: Dr. Rembrandt Wolpert  
Department of Music

### Abstract:

*This article presents part of my research on a type of electronic music known as Japanese noise carried out for my Honors Thesis in Music, Creating Silence through Noise: an Aesthetic Approach to the Sounds of "Japanese noise". It introduces Japanese noise, its origins in the 1970s and 1980s, its musical influences, and the early distribution of its pieces. The first generation of Japanese noise artists and their perceptions of Tokyo are then discussed. Finally, the possibility is advanced of a correlation between such perceptions of the city and the sounds of Japanese noise.*

### Japanese noise:

Around the late 1970s and early 1980s a small group of young Japanese people started to experiment with the synthesis and recording of sound<sup>1</sup>. They were certainly not the first to do this in Japan; ever since the opening of the Electronic Music Studio at the Tokyo National University of Fine Arts and Music in 1966 (Shimazu, 1994, p. 104), and the subsequent inauguration of other electronic music studios at universities throughout the country, these processes had been everyday fare. The novelty was that these young Japanese, pioneers of a music genre nowadays known as *Japanese noise* (or *Japa-noise*), were manipulating sounds from home. Moreover, by way of sound recording, processing, sequencing, and sound synthesis, they were organizing sounds and assembling them together to finally turn them into what they called "pure harsh noise".

The coining of the term "Japanese noise" has been attributed to the Japanese composer Akita Masami a.k.a. Merzbow who is said to have used it already in 1979 (Novak 1999a, p. 23, Caspari and Manzenreiter 2003, p. 64). Nowadays the term is used internationally. It groups together diverse Japanese soloists and bands whose music is directly associated with genres such as punk, metal, electronica, free-jazz, and pop, as I have been able to observe in some music stores in the Japanese cities of Tokyo, Hiroshima, Kyoto, and Ageo, as well as in Chicago, Denver, and New Orleans in the United States. The diversity of genres

currently embraced by the term Japanese noise can also be evidenced in Internet pages dedicated to it, such as <http://www.alchemy.cc/> (Alchemy Records) and <http://www.asahi-net.or.jp/ER6G-ITU/index.htm> (Noisebrybo: Guide to Noise World), among many others<sup>2</sup>.

Defining Japanese noise in clear and straightforward terms is not possible, as it has been used from 1979 onwards (Caspari and Manzenreiter, 2003, p. 64) either to avoid any other existing classification on the part of both artists and independent record labels or as a way to classify music that does not fit the parameters of classification already existent within music distribution networks (like Internet and music shops). However, I can say that, according to the primary and secondary sources I have been using for my research on Japanese noise, the majority of soloists/bands that have been classified since 1979 as Japanese noise play music which can be directly classified in at least one of the following genres: punk, rock, metal, pop, free-jazz, and electronica.

In spite of the present use of noise as an international taxonomy to indicate (outside Japan) either a music genre or a music style, the first generation of Japanese noise artists did not "see themselves as producing a distinctive format but as explorers searching for new sounds and ways of expression" (Caspari and Manzenreiter, 2003, p. 64). Japanese noise originated simultaneously but independently in the cities of Tokyo and Osaka. In the latter city, Japanese noise comprised predominantly punk and hardcore bands using electric guitar(s), voice, bass, and drums. In Tokyo, it consisted of solo projects, such as Merzbow, playing with cassette tapes, tape loops, distorted broken instruments (e.g. broken electric guitars), effect boxes, mixers, analogue synthesizers, contact microphones, and other inexpensive sound-machines. From now on in this paper I shall concentrate on the first generation of "Japanese noise made in Tokyo"; and I shall refer to it, indistinctively, as Japanese noise (or Japa-noise) or as first generation Japanese noise.

The first generation of Japanese noise artists, such as Merzbow, Hijokaidan, Nakajima Akihumi, KK.Null, and Tano

Koji, were inclined towards the utilization of noise as a plastic material. By this I mean that they tended to employ noise as a block of sonic matter that could be molded, shaped, and given form through composition. This plasticity of noise was due in part to the type of audio technology available at the time (see Fig. 1) but in part also - and more importantly - to the training some of the noise artists had had in different fields of the arts. Merzbow, for instance, studied Painting and Art Theory at Tamagawa University; Nakajima Akihumi studied Industrial Design at the Kyoto University of Art. Both Merzbow and Nakajima still work in the fields they originally studied, and they both have explained that when working with sound they do not do so as music composers, but rather as a plastic artist and an industrial designer, respectively (Hargus, 1997; Sjögren, 2001). Furthermore, each of these artists has expressed his use of noise consistently with his particular artistic background: Merzbow has described it in terms of the "colorful inks of an illustrator" (Merzbow interviewed by Hargus (1997)); Nakajima, viewing it as an aspect of a given object, exploring such an object as something visual, tactile, and physical, as well as a sound-source, has stated: "I'm always interested in both the sound itself and the image from the source. I want to keep each release united with the sound-source as much as I possibly can." (Nakajima interviewed by Duguid (1998))

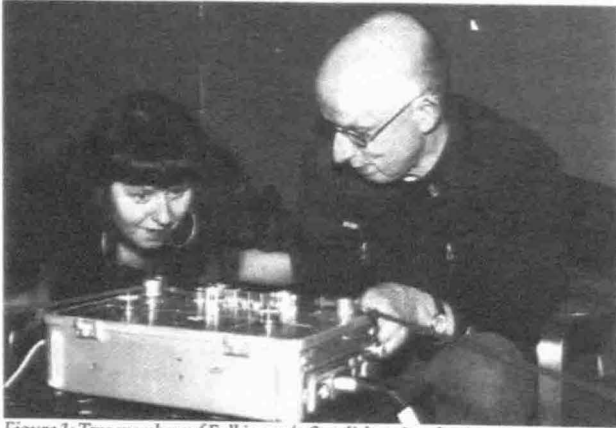


Figure 1: Two members of Fylkingen (a Swedish society for the promotion of new music and intermedia art) making a tape loop on a 1960s reel-to-reel machine<sup>3</sup>. Making a tape loop is a manual activity. Sound is treated as a physical thing (i.e. as tape) that can be cut/pasted and manipulated manually in many different ways.

### Influences:

Both free jazz (from Japan and the USA) and German Krautrock (German rock from the 1970s)<sup>4</sup> were among the main influences on Japa-noise: with free jazz it had in common the experimentation with timbre, freedom of form and tempo, the emphasis on improvisation, and the thirst for innovation; and with German Krautrock it shared the do-it-yourself ethos (DYS), the use of commercial audio-equipment, the extended use of analog synthesis, and a deep interest for surrealism. However, whereas German Krautrock had been directly inspired by the

German electronic avant-garde, and very specially by the music experiments of Karlheinz Stockhausen (Scaruffi, 2002), Japanese noise remained completely disconnected from the Japanese electronic musicians who, since the 1950s, had been composing at university studios throughout Japan. (And this was so although Japa-noise artists were actually familiar with the music of Stockhausen (Fig. 2) and of other European composers such as Pierre Henry, Luc Ferrari, and Iannis Xenakis.)

Indeed, the first group of Japanese noisicians did not even have contact among themselves: they started playing with noise as separated individuals, experimenting with it for their own pleasure, while recording and distributing their own cassette-tapes from home.

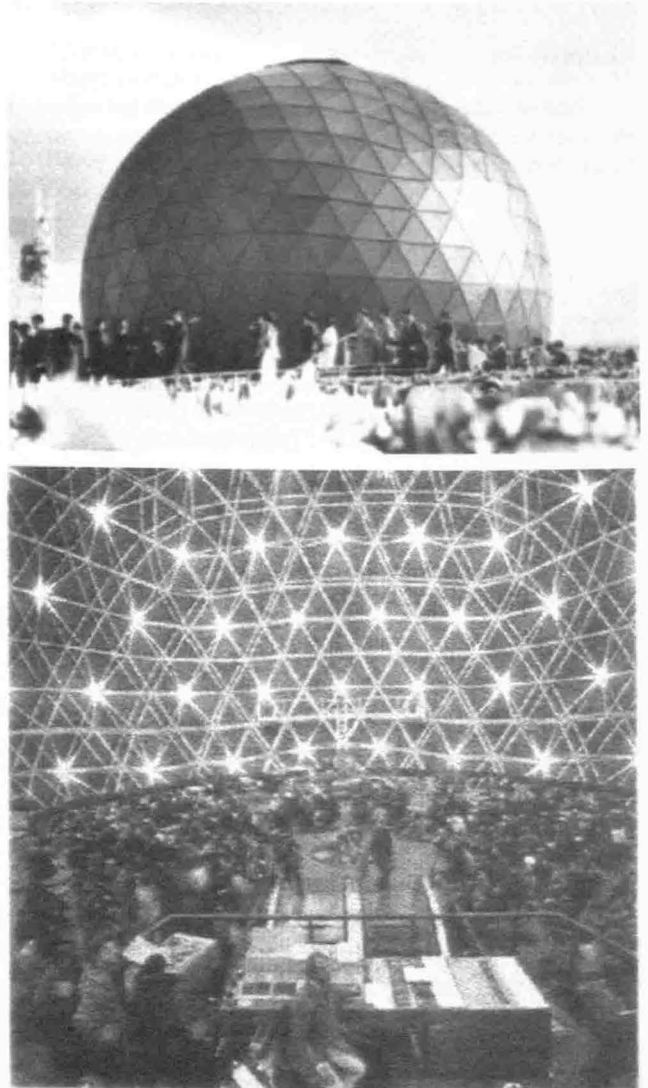


Figure 2: Evidence of Karlheinz Stockhausen's presence in Japan: This special concert hall based on Stockhausen ideas was built by Germany for the 1970 World Expo in Osaka, Japan. For 180 days music by different German composers was played through an innovative surround-sound system<sup>5</sup>.



### Japanese noise: early distribution:

Japanese noise started, then, as a homemade sonic experiment. It remained like that until the end of the second half of the 1980s when Japa-noise artists began to perform for audiences constituted of more people than merely themselves. What motivated the change, or why was it synchronic? These are questions to which answers are still blurred. Nevertheless, a partial explanation can be found in the wide acceptance of Merzbow's sound-works by numerous Russian, Western European, and USA underground music circles after he successfully toured these regions in 1988, 1989, and 1990, respectively. Nakajima Akihumi has confirmed this in a recent interview:

I think [that] since the mid 80s Japanese noise music became popular because of the many appearances by Merzbow or Hijokaidan who were already using noise on record. At the time, any foreigners were surprised that it [Japa-noise] was so loud and noisy. (Nakajima interviewed by Sjögren (2001))

Merzbow's international success attracted the attention of other Japanese artists and musicians who were working at the time on similar or related sound-pieces on their own, for themselves. And so noisicians were finally able to set up communication channels through which they started to meet, to listen to each other's sonic experiments, to collaborate, and to play together.

During the first half of the 1980s when Japanese noise was still in a cocoon state, the distribution of the tapes where most of the noise-works were recorded was practically null. A major exception to this collective introversion was Merzbow who found himself participating for several years in an international chain of mail art:

When I started Merzbow the idea was to make cheap cassettes which could also be fetish objects. I recorded them very cheaply and then packaged them with pornography. I got very involved in the mail art network which included home tapers like Maurizio Bianchi, Jupiter Larser [sic]<sup>6</sup> of Haters, and Trax of Italy. (Merzbow interviewed by Hensley (2006)).

Merzbow's first artistic interchanges, carried out by means of the postal service, were significant in different ways. First of all, I see the connections Merzbow established through this process of art-posting as essential for his later touring success; Merzbow mail exchanges thus played an important role in the "coming out" of noise. Second, the use of postal services for the distribution of noise-tapes was perpetrated over time, although the mail network was later replaced by the Internet. According to Caspari and Manzenreiter, the channels chosen for the circulation of Japa-noise (i.e. the postal services and then the Internet) gave the genre an anti-commercial outlook, and played an important role in transforming Japanese noise from "subculture

to cyberculture" (Caspari and Manzenreiter, 2003). Last, Merzbow's utilization of pornography in conjunction with his noise-tapes was (and it still is) misinterpreted outside Japan. In due course, Japanese noise has come to be wrongly linked with a type of commercial, sexual imagery that has nothing to do with the Japa-noise artists' intention. From my perspective, this false association has resulted from the contrast between the normality of most Japanese noise artists and the deviant, violent, extravagant individuals they have been expected to be by those outside Japan (see Fig. 3).

In other words, non-Japanese audiences of Japa-noise<sup>7</sup> have expected the harshness, roughness, and shrillness that characterize the first generation of sound-works to materialize into objects. Hence, the early pornographic collages and later bondage images utilized by Merzbow with his noise-releases have been trapped in the midst of these assumptions. Both pornography and bondage pictures have been put randomly on the covers of different Japa-noise releases issued outside Japan, without knowledge of the images employed, and lacking any connection between the images and the sounds they are supposed to portray. Merzbow, a plastic artist aware of the weight and importance of images, has referred to this attitude:

All bondage pictures I use are taken by myself. I know who the models are (...). I know the exact meaning of these pictures. This is very different from people using (...) images from Japanese magazines. I know that there are many bondage images associated with Merzbow releases. But many of these releases use stupid images without my permission. (...) I don't like the easy idea of using images without the knowledge of the image itself. So it's meaningless to create ideology by using pornography without the correct knowledge of the image itself. (Merzbow interviewed by Hensley (2006)).



Figure 3: Front cover of Akita Masami's vegetarian cookbook *My Vegetarian Life*. A kind way of living for me, the animals, and the earth. Both Akita's vegetarianism and his active participation in animal rights movements (he is an activist of peta) contrast with the deviant images that people fit him into, due to the intensity of the sound of his noise-works.

**Japanese noise, volume, and performance:**

When Japa-noise went “on stage” around the mid 1980s, it did so with sound volume - a volume so loud that in the bars hosting Japanese noise performances, no conversation could have been held, no polite *sumimasen*<sup>8</sup> heard by the waitresses and waiters. The sound-works were played so loudly that they imposed a vow of silence upon their listeners, almost all Japanese noise artists themselves at that time. To date, I have been unable to locate any particular accounts of whether Japanese noise artists had difficulties finding performance spaces due to this loud volume, or of whether their performances had to be adapted in order to comply with the anti-noise and vibration pollution regulations effective in Tokyo since the 1970s<sup>9</sup>. However, a comment by Merzbow confirming that “in the early days (. . .) people thought that the music [Merzbow’s] was just too difficult and loud” but “now [1999] grindcore and techno people come to see Merzbow,” and therefore, “(. . .) we [Merzbow and his collaborators] have been getting more places allowing a performance than ever before” (Hensley, 2006), could perhaps indicate that at least until 1999, finding a venue for playing Japa-noise inland was a difficult task for Japanese noise artists.

Perhaps the performance of Japanese noise in high volume has not only been a matter associated with the aesthetics of Japanese noise itself, but it may also be a resource through which Japanese noise artists have attempted to keep their noise-pieces out of the reach of the curious and the many. Kyoto-based noise artist Nakajima Akihumi (N), interviewed in 2001 by Klas Sjögren (S), has said in this respect:

S. When I first heard your music it felt like a fierce blow to the stomach. Have people ever been physically sick while you were performing harsh noise?

N. Yes maybe. . . probably me too (laughs). I got the same feelings when I first listened to noise music. I was fascinated by the noise because I never had had such an experience before and so I think that’s one of the most interesting points of noise.

S. The physical pain?

N. Yes. It also has pain or fear, but after listening to a very harsh sound you get a very comfortable silence. After listening to noise any small sound gets interesting.

S. So pain is good?

N. (Laughs) Yes, pain is good afterwards.

Although some people would agree that pain may not be the best possible experience one can go through by listening to Japanese noise, it is definitely an important aspect of Japanese noise performances. Both volume and pain can make Japa-noise performances into very physical events. They can induce Japanese noise audiences to simultaneously hear and feel the sounds of Japanese noise, as these sounds vibrate in tandem throughout

people’s bodies and ears. Merzbow’s descriptions of noise as something erotic, intimate, and unconscious (Hargus, 1997; Hensley, 2006) are connected with this volume-induced physicality. Loudly played noise is erotic in that it touches the body, its vibrations are in permanent contact with people’s skin. Loudly played noise is intimate in that it induces audiences to be silent, quietly immersed in themselves. Finally, loud noise is unconscious in that it runs unaware, free of form and pitch, causing pain to some and getting others into trance states (Hensley, 2006).

**Japanese noise artists perceptions of Tokyo:**

It does not take too long to write it down: Japanese noise artists have thought poorly of Tokyo. Their perception, I believe, echoes the voices of many other Japanese for whom Tokyo is just space, not place, for whom Tokyo is a point of encounter in permanent transition. Referring to Kobayashi Hideo’s essay *Kokyoo o ushinatta bungaku* “Literature of the lost home”, Seiji Lippit has touched upon this point, saying that “[Tokyo] serves not as a repository of accumulated memories - the necessary condition for a home to function as such - but only as an ever shifting marker of disassociation from the past” (Lippit, 2002, p. 3). More specifically, Japanese noise artists have characterized Tokyo as one of the more crowded and noisier cities in the world. This latter aspect, Tokyo’s noise, has been mentioned by both Merzbow and Nakajima Akihumi for influencing the sounds of their own Japa-noise-works. In an interview with Chad Hensley (H)<sup>10</sup>, Merzbow (M) has observed:

H: How has growing up in Japan affected your Noise creation?

M: Sometimes, I would like to kill the much too noisy Japanese by my own Noise. The effects of Japanese culture are too much noise everywhere. I want to make silence by my Noise. Maybe that is a fascist way of using sound (Hensley, 2006, interview conducted in 1999).

Moreover, Merzbow, interviewed by David Novak, has added:

M: I think that Japan, especially Tokyo, is not beautiful. Very ugly, no beauty. Western cities have more tradition. But Tokyo is just ugly, out of control. People still don’t care about destruction (Novak, 1999b, p. 28).

Nakajima Akihumi (N), interviewed by Klas Sjögren (S), has commented in a similar direction:

S: Do you think that the development of this Noise wave from Japan has a connection with Japan as a country or Japanese people . . . that the music has been so hard and noisy?

N: Japanese culture is very different from European or American culture. As you know Japan is very small

but very modern and contemporary. Tokyo is the most crowded and busy city in the world and always noisy. There are the noisy sounds from the streets or from between the buildings. In some ways part of our lives are always noisy (Sjögren, 2001).

And, Nakajima Akihumi (N) interviewed earlier by Brian Duguid (D) had also said:

N: Noise music is a kind of music from the city. It's always as noisy as any modern city, although not as much as Tokyo or Osaka (Duguid, 1998).

On several occasions, in different interviews posted on electronic popular-magazines and other e-spaces on the Internet, Japanese noise artists have denied that their music has any relation whatsoever with Japanese music. (In reality they are usually asked if their music has connections with Japanese traditional music, and the traditional bit of the expression really infuriates them: it is incomprehensible to them.) Furthermore, with the exception of Nakajima Akihumi who has stated that he feels "something in common with these other artists [Japanese noise artists], as we make similar sounds, even though the methods and concepts we use are different from each other" (Nakajima interviewed by Duguid (1998)), several Japanese noisicians have repeatedly refused to be classified as such (i.e. as Japanese noise artists), asserting that they just do their music independently; and this is in spite of their own frequent allusions to a "first and second generation of Japanese noise". However, as we have just read, both Merzbow and Nakajima have conceded that if anything **Japanese** has actually influenced their sonic-making it is, generally speaking, the loudness of Japan, and more precisely, the noisiness of Tokyo.

Each of these artists has responded differently to their common perception of the city in terms of noise. To begin with, Merzbow's response has been explicit. His artistic intention has focused on canceling out the outer-noises of Tokyo with his own noise, in a way like portable-music-player users have tended to turn up the volume of their devices to drown out external environmental noise (e.g. street noise, unwanted music played in public spaces, annoying conversations...). Conversely, Nakajima Akihumi's response has been implicit. Unlike Merzbow, Nakajima's perception of Tokyo has not prompted him to take a tangible artistic action; neither has it taken him through a specific artistic path. Nonetheless, Nakajima's discernment of the soundscape<sup>11</sup> of Tokyo as noise is latently present in his sound-works, and I believe it has influenced his overall artistic creation in terms of style, directing it (the style) towards Japanese noise.

The above interviews with Merzbow and Nakajima Akihumi are the only sources I have been able to find in which Japanese noise artists have concretely mentioned their personal perceptions of Japan and/or Tokyo. In spite of this apparent informative gap, I nevertheless think that it can be safely stated that the majority of Japanese noise artists - first generation - have thought of Tokyo as a noisy, busy, crowded city. This claim may be further

supported by the data on noise in Tokyo and on protest and legislation concerning noise and vibration pollution in that city which I present and analyze in another paper (Alarcón Jiménez, 2006b). And while, with the exception of Merzbow and Nakajima Akihumi, we may not be able to know the particular responses of Japanese noise artists to their impressions of Tokyo's noise, we can infer that such impressions are not only present in these artists' sound-works but are also correlated with their characterization as (the first generation of) Japanese noise.

#### In conclusion (Dreaming noise):

Somehow, the sounds of the neighborhood seem changed. What had resembled the rippled hissing of late-night TV static, this morning sounds like a thousand pipe organs in massed unison, all pulled as per a master score, now in chorus, now a capella, creating a great symphony. Countless buildings visible from my balcony are resonating, perhaps with a melody unique to each. (Shimada, 2002, p. 431)

The quote above is the closing paragraph of Japanese writer Shimada Masahiko's short story, "Desert Dolphin". The context in which this quote appears, and its meaning in the story, are as follows: an angel who ten years before was condemned to live on earth meets, coincidentally, a younger angel who has just fallen down from heaven. The experienced angel invites the newcomer to spend the night at his place (an apartment in Tokyo) in order to teach him and give him advice on how to live on earth. On their way to the apartment at night, the fallen angels walk together through the streets of Tokyo. Suddenly, the experienced angel, also the narrator of the story, comments:

Outside, the sound of bottles smashing, the whirr of taxis racing in low gear along narrow streets, the chorus of electronic beeps pouring from a video arcade, the laughter of a group of pub-crawlers, everyone making noise, everyone a street urchin making sounds nobody cares to listen to...

And then this exchange develops:

'Noisy place, isn't it, the world?'

'Not necessarily. It was pretty much the same in Heaven.'

'I wonder. Surely Heaven has a better sense of harmony, nothing near the random racket of Earth.'

'My ears must be deceiving me. These earthly sounds of yours are like beautiful chords.'

'The din of the streets a beautiful chord? He must be hearing the sounds of earth with his heavenly apparatus intact, because in such ears even noise sounds like harmony' (Shimada, 2002, p. 420).

As the story continues, the angels arrive at the apartment, and they talk, sharing their life stories. Late at night, before falling asleep, the experienced angel confesses to the other that

he wished heaven would come down to earth. The story finishes as this angel wakes up the next morning feeling something strange, something different, and it is then that he says: "Somehow, the sounds of the neighborhood seem changed. . ." (Shimada, 2002, p. 431)

With my understanding "music" as something that "is rather a paradigm of a culture in a historical context; a complex signal, gratifying to the hearer for cultural reasons" (Picken, 1999, p. 9), the story above illustrates, for me, the sense and aesthetic direction of Japanese noise. To begin with, the way in which the Japanese noise artists relate to the soundscape of Tokyo before and after turning its sounds into their own sonic pieces is similar to the way in which the experienced fallen angel relates to the sounds of this same city before and after falling asleep. All of them, the angel and the Japanese noise artists, have a feeling of dislike towards the noises and the crowds of Tokyo. But they are all able to transform these sounds in such a way that the final sonic outcome is aesthetically enriching. In the middle of Shimada's story, the experienced angel comments to the newcomer that "music, and dreaming, are the ties that bind Heaven and Earth" (Shimada, 2002, p. 422). Amazingly enough, this statement fits perfectly to exemplify the paths followed by the Japanese noise artists, on the one hand, and by the angel, on the other, towards hearing the unwanted noises of a city as beautified noise. For, just as the angel dreams and wakes up hearing the sounds of earth as if he were in heaven, the Japanese noise artists capture, transform, and organize the "din of the streets" (Shimada, 2002, p. 420) of Tokyo, by making music with them, by turning them into their own noise.

### Acknowledgements:

I am deeply indebted to Dr. Elizabeth J. Markham and Dr. Rembrandt F. Wolpert for their invaluable help and precious advice. They have been extremely supportive in the writing and editing of both this article and my Honors Thesis in Music, *Creating silence through noise: an aesthetic approach to the sounds of "Japanese noise"*. I would also like to thank Dr. Wolpert and Dr. Markham for all their encouragement for the presentation of my paper, "Making silence through noise: the soundscape of the Tokyo of the 1970s", at the conference *Music in the Public Sphere*, May 12-13, 2006, University of California, Los Angeles (ucla). A third aspect of my thesis that indirectly supports this paper involves music analysis of selected Japanese works. This work was carried out (during summer-study abroad) under the supervision of Dr. Leigh Landy at DeMontford University, Leicester, England, and at CCMIX, Paris, France. I wish to thank all involved, especially Dr. Landy.

### Endnotes:

<sup>1</sup>In a conference paper (Alarcón Jiménez, 2006b), also drawn from my Honors Thesis, I suggest that the emergence of the phenomenon of Japanese noise at this time should be considered in the light of the strong, and even violent, public protest against (and governmental legislative reaction to) noise and

vibration pollution in a rapidly developing Tokyo.

<sup>2</sup>The analytical and careful reading of Japanese noise CD covers, interviews with noise artists on the Internet, artists' web-pages, Internet web-pages (specially those of record labels), concert pamphlets, related articles, and David Novak's (1999b) dissertation on Japanese noise, together with personal visits to music stores in both Japan and the United States, have led me to conclude that the process by which an artist's music becomes classified as Japanese noise usually takes place through one or more of the three following steps: (1) the artists themselves categorize their music as Japanese noise (Merzbow, CCCC, Boredomes, Melt Banana, and Hijokaidan), (2) the artists record with a label which has depicted itself as a noise label (i.e. Japanese labels Alchemy, MSBR), (3) when a Japanese artist's music can not be classified in any of the genres already established by popular music distributors on the Internet and/or CD shops, the music gets classified as Japanese noise in so far as it is both composed and performed by Japanese people, and as long as the timbre material of the music has something unusual or "noisy" in comparison with regular popular music standards (Cybo Matto, Ryoji Ikeda, and Otomo Yoshide).

<sup>3</sup><http://home.swipnet.se/sonoloco12/fylkingenevents/loop.html>

<sup>4</sup>Term taken from Ulrich D. Einbrodt's "Space, Mysticism, Romantic Sequencing, and the Widening of Form in German Krautrock during the 70s". It includes bands like Amon Düel, Can, New!, Guru-Guru, Kraftwerk, Cluster, Tangerine Dream, Ash Ra, Faust, and Popol Vuh. (Einbrodt, 2001)

<sup>5</sup><http://www.medienkunstnetz.de/works/stockhausen-im-kugelauditorium>

<sup>6</sup>The name should read Larsen.

<sup>7</sup>For sound-examples, see <http://www.steinklang-records.at/koji-tano.html>

<sup>8</sup>"Excuse me" or "I am sorry". In Japan this word is usually used by customers of Japanese bars and restaurants to call the attention of waitresses or waiters.

<sup>9</sup>I address these regulations in detail in another paper (Alarcón Jiménez, 2006b).

<sup>10</sup>This interview is source of inspiration for the title of my Honors thesis (Alarcón Jiménez, 2006a).

<sup>11</sup>"Soundscape" has been defined by Raymond Murray Schafer as "the sonic environment" (Schafer, 1977, p. 247)

### References:

- Alarcón Jiménez, A. M. (2006a). *Making Silence through Noise: An Aesthetic Approach to the Sounds of "Japanese Noise"*. Honors Thesis, University of Arkansas.
- Alarcón Jiménez, A. M. (2006b). Making silence through noise: the sound-scape of the Tokyo of the 1970s. *Music and the Public Sphere*. Conference paper at the University of California, Los Angeles.
- Caspari, C. and Manzenreiter, W. (2003). From subculture to cyberculture? the Japanese Noise Alliance and the internet. In Gottlieb, N. and McLelland, M., editors, *Japanese Cyberculture*. Routledge, New York, NY.
- Duguid, B. (1998). Aube. *EST Magazine*. <http://media.hyperreal.org/zines/est/intervs/aube.html>.
- Einbrodt, U. D. (2001). Space, mysticism, romantic music, sequencing, and the widening of form in German "Krautrock" during the 70's. Web publication, Universit at Giessen. <http://geb.unigiessen.de/geb/volltexte/2001/592/pdf/p010004.pdf>.
- Fletcher, P. (1999). *World Musics in Context -A Comprehensive Survey of the World's Major Musical Cultures*. Oxford University Press, Oxford.
- Hargus, B. (1997). Merzbow. *Perfect Sound Forever*. <http://www.perfectsoundforever.com/>

[www.furious.com/perfect/mmerzbow.html](http://www.furious.com/perfect/mmerzbow.html).

Hensley, C. (2006). The beauty of noise. *EsoTerra*, 8. <http://www.esoterra.org/merzbow.htm>.

Lippit, S. M. (2002). *Topographies of Japanese Modernism*. Columbia University Press, New York, NY.

Novak, D. (1999a). Music and language. Unpublished paper.

Novak, D. (1999b). The national and the transnational in the Japanese underground. Master's thesis, Wesleyan University.

Picken, L. E. R. (1999). Foreword. In Fletcher (1999).

Scaruffi, P. (2002). *A History of Rock Music*. Web publication (book), <http://www.scaruffi.com/history/long.html>.

Schafer, R. M. (1977). *The Tuning of the World*. Knopf, New York, NY.

Shimada, M. (2002). Desert Dolphin. In Goossen, T., editor, *The Oxford Book of Japanese Short Stories*, pages 417-431. Oxford University Press, Oxford.

Shimazu, T. (1994). The history and computer music in Japan: Significant composers and their works. *Leonardo Music Journal*, 4:102-106.

Sjögren, K. (2001). Aube. *Time Stereo* 2001. <http://www.timestereo.com/aube/index.shtml>.

### Faculty comments:

Dr. Rembrandt Wolpert, Ms. Alarcón Jiménez's mentor had the following comments about his student's work:

Ana Maria's study of the powerful and perhaps even shocking electronic music phenomenon of *Japanese noise* (also *Japa-noise*) for her Honors Thesis in what I know as Systematic Musicology has taken me along - over at least two years - on one of the most innovative, imaginative, and surprising paths in my (now longish) career of supervising undergraduate debuts in research. This article, on the genesis of *noise* and the aesthetics of the pioneer Tokyo-based *noisicians*, draws from her thesis *Creating silence through noise: an aesthetic approach to "Japanese noise"*, and reflects one of the three integrated aspects of her investigation. She lays out here her argument that Japa-noise artists' perceptions of the noisiness of the city (Tokyo) are canceled out, or transformed, through their making of their own aesthetically enriching *noise*. It is probably useful to state that, in her full study, this proposition is etched out against the background of the Japanese relationship to nature and the environment, manifest in the arts as a particular aesthetic, and is sized up and contextualized alongside a consideration of public outcry and government legislation during the 70s and 80s over noise and vibration pollution in a Tokyo under construction as it was then. And the third, technical prop to the thesis is a visual-imagery-based

musical analysis of representative *Japa-noise* works that aims, by attempting to represent in temporal and acoustic space the visual metaphor and plasticity of form articulated by the noise composers themselves, to support her position that their *noise*, explicitly or implicitly, relates to the noisiness of the city.

Embedded in Japanese aesthetics (disturbed by noise and vibration), then, this study of what is often characterized as chaotic and unstructured, named by the artists themselves as "pure harsh music", and typically finds its followers internationally in underground music scenes, is of remarkable refinement and breadth. Ana Maria has worked through scholarly literature on East Asian concepts of sounds and soundscapes beginning with a Chinese dictionary of the first century CE, on through aesthetic treatises on Japanese *Noh* drama, studies of Japanese architecture, town planning, and gardening, government white papers and statistics on noise and vibration pollution, ecological approaches to music perception and musical meaning, and out onto the streets and into the CD shops of modern Japan and the Americas in order to build her carefully painted, carefully worded picture of *Japanese noise*. She has a long-standing interest in composing electronic music herself, experience which let her spend profitable summer-abroad research on her thesis-analysis of Japa-noise in the electronic music studios of the *Music, Technology and Innovation Research Centre* at De Montfort University, UK (with my colleague Dr. Leigh Landy) and, as well, at the *Centre de creation musicale Iannis Xenakis (CCMIX)*, Paris (as participant in a Summer School).

I am delighted with the approach, openness, and originality of Ana Maria's study, and consider myself qualified to appreciate what it means to write an excellent, first piece of research in a foreign language. That a second paper drawn from the thesis *Making silence through noise: the soundscape of the Tokyo of the 1970s* was accepted for *Music and the Public Sphere: A Conference Presented by Echo: a Music-Centered Journal*, May 12 and 13, 2006, Los Angeles, CA, is just a first stage in recognition for this particular work, and only a first stage too in what I anticipate will be very fine academic career.

## **SECTION II: SOCIAL SCIENCES AND BUSINESS**

**ACCOUNTING, FINANCE, AND INFORMATION SYSTEMS**



## SEC ENFORCEMENT ACTIONS: AN ANALYSIS OF ECONOMIC DETERMINANTS OF RESTATEMENTS

By Rivka Berman  
Department of Accounting

Advisor: Dr. Don W. Finn  
Department of Accounting

### Abstract:

*This paper investigates the economic determinants of restatements, focusing on companies that were the object of an SEC Enforcement Action. A sample of 30 restatement firms is matched with 30 non-restatement firms in the same industry and same business and with a similar size as measured by assets. It was found that the use of a Big 5 audit firm reduces the incidence of restatement. More specifically, use of Ernst & Young as a Big 5 audit firm significantly reduces the possibility of restatement. Variables testing debt to equity ratio, proportion of management ownership and proportion of blockholder ownership as determinants of restatement were not significant with this particular sample. Because of the strong increase in the number of restatements in recent years, it is important for a company to reduce the potential probability of restatement of its financial statements. This can be done through the utilization of a Big 5 audit firm.*

### Introduction:

When viewing a set of financial statements, many would assume that the statements truthfully represent financial results for the company and that these results are free from fraudulent information. This conclusion is partially due to the fact that statements are analyzed by investors, analysts, and competitors who use the information contained in financial statements to make decisions about the company. In recent years, it appears that the assumption of truth may not be completely true, judging from accounting scandals that have dominated popular press headlines for several years. We have learned from these scandals that financial statements for many companies contained fraudulent information which was used by analysts and investors to make judgments about the future profitability of those firms. By the time fraudulent information was discovered the information backing their decisions was misleading, many investments had already lost a substantial portion of their previous value.

After a portion of the information in financial statements is determined to be incorrect, the company must correct the information and file a restatement with the Securities and Exchange Commission (SEC). Restatements include financial statements that were subject to changes in GAAP (such as the switch from FIFO to LIFO inventory methods), subsequent events (such as stock splits, mergers, and divestitures), and true fraudulent reporting, in which information is materially misrepresented. When fraudulent actions are suspected, the SEC often conducts an investigation into the matter and may issue an Accounting and Auditing Enforcement Action, often called an Enforcement Action. This document outlines the fraudulent actions, providing background about the company, the amount of restatement, and nature of restatement, and any future actions taken against the company or its management. In essence, an Enforcement Action publicly identifies the company that participated in fraudulent wrongdoing.

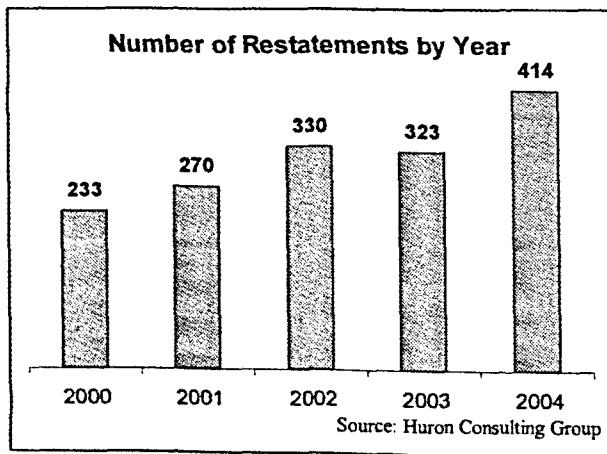
This paper investigates the economic determinants of restatements. It will focus only on companies who have restated earnings and have been the target of an Enforcement Action. This will highlight companies that have had fraudulent actions, rather than other less serious causes of restatements. To discern whether an accounting action is actually an economic determinant of restatement, all restatement firms are matched (a control group) with an equivalent non-restatement firm. The potential economic determinants between restatement and non-restatement companies will be measured using the leverage, proportion of management ownership, and proportion of blockholder ownership.

After analyzing a sample of 30 restatement companies and their matched non-restatement companies, it was determined that those firms audited by Big 5 firms were less likely to produce a restatement. In a second regression analysis, Ernst & Young was found to be the Big 5 firm whose clients were least likely to have a restatement. Variables used to test other hypotheses were not significant in this test possibly because of the limited sample available.

The remainder of the paper will be organized as follows. Section II discusses relevant background information about the occurrence of restatements. Section III reviews previous findings of studies that have discussed restatements. Section IV defines the sample and data collection methods. Section V discusses results of analysis of the relationship between restatement and non-restatement companies, and Section VI provides final conclusions.

### Background Information:

To understand the importance of restatements, one high profile example is the case of Enron. When Enron's restatement reduced net income by \$569 million for the years ended December 31, 1999-2000 and the first two quarters of 2001, the restatement was not just a change in numbers for a few accounts (Akhigbe, Kudla, & Madura, 2005). It also brought a series of changes that have had long lasting effects on overall investor confidence. Investors found it difficult to trust the accuracy and reliability of other companies when their Enron stock had lost nearly all of its value in a period of days. If a fraud as significant as that of Enron could occur within the sight of auditors, then investors may begin to wonder about the dependability of information provided by other public companies.



The incidence of restatements has become much more common since the year 2000, with a 22% increase from 2000-2004 according to data collected by the Huron Consulting Group. The data also shows a 28% increase in restatements between 2003 and 2004. Jeff Szafran, managing director of Huron Consulting, suggests that this increase might be attributed to the "unprecedented level of regulatory and audit scrutiny, driven primarily by the Sarbanes-Oxley Act of 2002" (Bryan, Lilien, Ruland, & Sinnett, 2005). Among its many provisions, the act requires CEOs and CFOs to certify the accuracy of financial statements, including a certification of internal controls, establishes an independent audit committee, and provides conditions of auditor independence. Szafran also points out the following conditions regarding the increase in restatements:

Public companies spent significant amounts of time and money to comply with the requirements of Sarbanes-Oxley Section 404, and may have found some mistakes in the process. (This section requires that management attest to responsibility and accuracy of internal controls.)

The Public Company Accounting Oversight Board (PCAOB) began reviewing the audit practices of the major accounting firms.

The SEC budget jumped to almost twice what it was during 2001, and it used that budget to hire more professionals to enforce the law.

The SEC established a new Office of Risk Assessment to look into the practices of certain industries.

Auditors were doing more work, including testing companies' internal controls. (Bryan et. al., 2005)

It is important to also note that the increase in earnings restatements has spiked within the last six to eight years and still reflects a small percentage of overall filings. A study by the FEI Research Foundation reviewed the incidence of earning restatements with a sample of 1080 restatements between 1977 and 2000. It found that the average number of restatements between 1977 and 2000 was only 49, with a large spike between 1998 and 2000 (Moriarty & Livingston, 2001). More importantly, the study found that the average number of restatements since 1995 represented 0.67% of the average number of reporting companies (Moriarty & Livingston, 2001). This would suggest that, while earnings restatements are significant events in the life of a company, the overall quality of financial statements still remains high.

A 2004 General Accounting Office study supports Szafran's analysis of conditions that led to an increase in restatements. It says about half of companies found their own mistakes, while external auditors found 2.5% of mistakes and other external parties found 9% of mistakes (Wallace, 2005). Another 4.5% were discovered by the SEC, the type of case that will be analyzed in this paper (Wallace, 2005). This breakdown shows that companies are usually forthcoming when they discover an error and may be the best defense against incorrect accounting practices.

The Huron Consulting report also discloses further trends in restatements. It found that of the 414 restatements in 2004, 253 (61%) restated annual reports, which are audited by external auditors (Bryan et. al., 2005). They also discovered that almost 40% of restated annual reports were multiple year restatements, signaling "flawed accounting policies" and long-standing errors instead of one time errors (Bryan, et. al., 2005).

Overall, the size and number of restatements appears to be increasing even with the negative repercussions experienced by the company following a restatement. Most mistakes that cause

restatements are found by the company itself and are likely to cover multiple years. Even with the overall increase in the numbers of restatements, less than 1% of companies file a restatement in a year, demonstrating the significance of a restatement in the life of a company

### Previous Findings:

The majority of prior literature discusses effects of restatements on the stock market. These studies overwhelmingly conclude that earnings restatements produced negative stock returns. In a sample of firms that restated earnings between 1976 and 1985, William Kinney and Linda McDaniel found that on average stocks earned negative returns between the release of false financial statements and the release of the correction (Kinney & McDaniel, 1989). The amount of negative return was quantified by Palmrose, Richardson, and Scholz. They found that the mean abnormal return from a sample of 403 restatements from 1995-1999 was -9.2% over a two day announcement period (Palmrose, Richardson, & Scholz, 2004). They further concluded that the average stock price change was even larger than -9.2% when restatements included an indication of management fraud, had large material dollar effects, and were initiated by auditors rather than the company itself.

Anderson and Yohn examined whether different causes of a restatement have larger effects on stock price changes. They found a larger negative reaction to revenue recognition restatements than other type of restatements (Anderson and Yohn, 2002). This demonstrates that revenue recognition affects investors' view of "firm value and information asymmetries" more than other types of restatements (Anderson and Yohn, 2002).

Most previous research centered on stock market effects from all types of shareholders. Hribar, Jenkins, and Wang, however, focused on institutional shareholders. This type of shareholder usually represents a blockholder (owns more than 5% of all shares outstanding), a group that will be examined further in this paper. The study provided three key conclusions about institutional shareholders. First, transient institutions, those focused on the short-term, predict earnings restatements one quarter prior to actual restatement, providing evidence that they have an information advantage over regular shareholders (Hribar, Jenkins, and Wang, 2004). This would be an important distinction to draw as this paper examines the relationship between blockholders and restatement companies. Next, the study found that institutional shareholders respond more negatively than other investors to the announcement of a restatement, having different interpretations of both the sign and weight of the restatement (Hribar et. al., 2004). Lastly, institutional holders are found to trade earlier than individual investors, who usually trade over a five day window around restatements (Hribar, et. al., 2004). Overall, the results of this study illustrate the importance of institutional shareholders to the market as a

whole because this type of shareholder can have an information advantage and is likely to respond sooner and more negatively than the typical shareholder.

Overall, prior research confirms that restatements generally cause the stock price of a firm to decline following a restatement, usually around 9%. Also, institutional shareholders have additional information backing their decisions, allowing them to make decisions earlier and more drastically than the average shareholder.

### Hypothesis Development

This paper is designed to test for the effect of economic determinants of restatements on financial statements. To do this, three groups that would have interest in seeing the best results for the company are examined. These are the debt holders, management, and institutional shareholders or blockholders. It is expected that each of these groups would exercise interest for the company to produce the best financial results possible. For example, debt holders expect a company to act in ways that ensure the future repayment of the debt. Management would seek to ensure continued employment and maximization of stock-based compensation. Blockholders would desire maximum returns on investments and would act in a manner to accomplish this objective. Hypotheses tested are as follows:

H<sub>1</sub>: There is a positive relationship between the debt to equity ratio and incidence of restatement.

H<sub>2</sub>: There is a positive relationship between the proportion of management ownership and the incidence of restatement.

H<sub>3</sub>: There is a positive relationship between the proportion of blockholder ownership and the incidence of restatement.

H<sub>4</sub>: There is a negative relationship between use of a Big 5 accounting firm and the incidence of restatement.

### Sample and Data Collection:

The sample used for analysis includes 30 companies that were the object of an SEC Enforcement Action regarding financial statements for the years 1999-2004. This information was obtained from a search of the selected Accounting and Auditing Enforcement Releases available on the SEC website. Each release was analyzed to see that it did include a restatement of quarterly or yearly financial statements (10-Q or 10-K). Releases were also analyzed to determine the quarter or year in which

restatement had the greatest effect on previously stated earnings. This period of time was then matched with a control firm that had not been the object of an Enforcement Action. Restatement and non-restatement companies were matched on three criteria: 1.) has the same North American Industrial Classification System (NAICS) code; 2.) a similar type of business as discussed in the "Business" section of each company's 10-K; 3.) the closest size as measured by total assets for the period restated.

The Compustat North America Industrial Annual and Quarterly databases were used to determine the total assets used to match companies. The databases were also used to pull the short-term debt, long-term debt, and total equity used to compose the Debt to Equity ratio used in analysis.

The proportion of management ownership and the proportion of blockholder ownership also used in analysis of economic determinants were gathered from the Proxy statements available on the SEC website. The Proxy for the period of restatement was used to gain this information. If a manager was also a blockholder, he or she was included in the management percentage and not the blockholder percentage. The blockholder percentage only represents those blockholders that were not included in the management percentage. This eliminates redundancy among the data collected.

For usage in the descriptive statistics section, the auditor at the period of restatement was also collected from the Proxy statement. Also, the Enforcement Releases were analyzed to determine common reasons for restatement which are displayed in the descriptive statistics.

#### *Descriptive Statistics*

Descriptive statistics of the sample of 30 restatement firms and matches are presented in tables 1-4. Table 1 presents the distribution of years represented by the restatements. In this sample, an equal number (11) of restatements concerned the years 2000 and 2001, while fewer restatements concerned later years. This represents the year affected by the restatement rather than the year the restatement was released. Likely, the need for restatements of information in later years has not yet been discovered or has not yet been the subject of an Enforcement Action by the SEC.

Table 2 represents the number of restatements by NAICS code for industry. It shows that nine of the 30 restatements (30%) used in the study came from the manufacturing industry. This includes sub-industry names such as engines and turbines, pharmaceutical preparations, special industry machinery, electric computers, and prefabricated metal buildings. Companies include Cummins, Inc., Bristol-Myers Squibb Company, DT Industries, Inc., and NCI Building Systems, Inc. Also notable are the six restatements from the information industry. This includes Time Warner, Inc. from motion picture and video tape production,

Gemstar-TV Guide International, Inc. from periodicals, and i2 Technologies, Inc. from prepackaged software.

Table 3 discloses the reasons for restatements found in the Enforcement Action document on the SEC website. Overwhelmingly, 12 of the 30 restatements (40%) involved improperly recognized revenue. This follows previous findings by a study sponsored by the FEI Research Foundation that also found the most prevalent reasons for restatement was revenue recognition followed by cost issues and loan loss provisions. In the sample used in this paper, cost issues (called improperly recorded expenses) and loan loss provisions (called treatment of losses) are also dominant issues.

Table 4 reviews the auditor responsible for the period of restatement for both the restatement firms and non-restatement firms. PricewaterhouseCoopers audited an overall majority of firms and a majority of the restatement firms also. Also notable, Arthur Anderson audited the second most number of firms, as the majority of time covered by the sample was before the firm was disbanded. Ernst & Young was largely more likely to audit a non-restatement firm rather than a restatement firm, while "other" firms were more likely to audit restatement firms. The "other" category includes firms that were not in the Big 5, such as regional and local firms. Even though these firms were more likely to audit a restatement firm rather than a non-restatement firm, most restatement firms had been audited by a Big 5 firm.

#### **Results:**

##### *Univariate Statistics*

Univariate statistics were used to examine the basic variation between restatement and matched (control) companies. Although these differences alone do not provide evidence to support individual hypotheses, they do provide insight into the sample used in this paper.

Table 5 presents the overall mean of variables from all causes of restatements and their matched companies. In this sample of 30 restatement companies and 30 matched companies, matched companies exhibited a higher asset to equity ratio, shown as 7.86 for matched companies and 4.46 for restatement companies. This illustrates that companies experiencing a restatement hold a lower amount of assets in relation to their equity, providing a possible explanation of their behavior. Also notable, restatement companies have a 2.6% higher percentage of blockholder ownership than matched companies. This may show that because restatement companies have a higher proportion of stock owned by these institutional shareholders, they may also be susceptible to the pressures from these blockholders. Such a type of owner would be more demanding than an individual stockholder because of the concentrated nature of ownership.

Table 6 presents the mean of only the two most common causes of restatement as previously discussed – improperly

recognized revenue and improperly recorded expenses. Of the firms that have improperly recognized revenue, restatement companies have a higher asset to equity ratio than matched companies. The mean of 2.24 for restatement firms is much lower than the mean of 4.46 from all causes of restatement. Also, assets to equity is higher for matched companies than restatement companies when reviewing the overall mean. Companies with improperly recognized revenues have a considerably higher percentage of blockholder ownership, representing the same situation as the overall mean of all restatements.

This ownership situation is negative when examining only the means of restatement firms and matched firms for improperly recorded expenses. For this situation, matched firms have a higher proportion of blockholder ownership than restatement firms. Restatement firms also have higher assets to equity, current debt to equity, long term debt to equity, and total debt to equity than matched companies. With the possible exception of assets to equity, this would follow conventional knowledge that the pressures of holding debt could lead to possible areas of fraud, therefore restatements

### *Regression Analysis*

The hypotheses were tested using two different regression models. In each model, restatement versus matched firms was used as the dependent variable, using a dichotomous indicator variable of 1 to indicate a restatement firm or 0 to indicate a matched firm. Other objective variables, such as Big 5 audit firm in Model 1 and individual Big 5 firm or smaller SEC audit firms in Model 2, respectively, were measured by using a 1 to signify use of the particular firm and a 0 to signify use of another firm. Also, the Log of assets was used to eliminate any size bias between firms with different levels of assets.

The models are as follows:

#### Model 1

$$\text{Restatement} = a + b_1(\text{assets}) + b_2(\text{debt/assets}) + b_3(\text{debt/equity}) + b_4(\% \text{ of management ownership}) + b_5(\% \text{ of blockholder ownership}) + b_6(\text{Big 5 audit firm})$$

#### Model 2

$$\text{Restatement} = a + b_1(\text{assets}) + b_2(\text{debt/assets}) + b_3(\text{debt/equity}) + b_4(\% \text{ of management ownership}) + b_5(\% \text{ of blockholder ownership}) + b_6(\text{EY}) + b_7(\text{Arthur Anderson}) + b_8(\text{Deloitte}) + b_9(\text{KPMG}) + b_{10}(\text{PWC})$$

Table 7 illustrates the mean, median, standard deviation, high and low values for each of the variables used in each model. It indicates the wide variation between the high and low values for the particular variable and gives a relative idea of the midpoint of the variable.

Table 8 presents the results of regression tests performed using the two models presented above. Model 1 shows that the

intercept and Big 5 audit firm results are significant at a 10% level. The p-value of the Big 5 audit firm variable can be divided in half to represent a one tailed test rather than a two tailed test. This is applicable because the influence of an audit firm is hypothesized to be positively related to fewer restatements, and thus it is a one-tailed test. In this case, it is that firms who use a Big 5 audit firm would experience fewer restatements, as large audit firms have been shown in previous studies to produce higher quality audit results. In their 1999 paper, Colbert and Murray present several reasons why major audit firms would produce a higher quality audit. These include taking advantage of economies of scale, guarding reputation, and utilization of human capital (Colbert and Murray, 1999).

The Big 5 audit firm variable is also significant because of its representation of the overall population of public companies. Firms who used an auditor other than a Big 5 firm received a "0" value for the Big 5 variable, and the proportion of firms using a non-Big 5 firm (10%) reflects the actual proportion of public firms using non-Big 5 audit firms.

The regression of Model 2 is also presented in Table 8. The intercept and Ernst & Young both produced significant results. At a 10% level in a one tailed test, firms audited by Ernst & Young are less likely (negative coefficient of -0.50) to audit a company that restated financial statements. This would follow the regression analysis of Model 1 that found companies audited by a Big 5 firm are less likely to produce a restatement.

The lack of overall significance in both models may be attributed to the small sample size. It is possible that with a larger sample size, results would be significant. However, an examination of the coefficients (although not significant) may provide some additional insight relative to the overall research question – what are the factors that contribute to restatements?

There is anecdotal information from the coefficients that may help one to interpret how managers might engage in mistatements of financial information. The most important variables are the percentage of management and blockholder ownership. In both models, negative coefficient of management ownership indicates that with increased percentages of management ownership, the probability of restatement decreases. A possible explanation of this is that management would likely want to avoid restatement to protect the value of its ownership in the company. A similar logical argument exists with the percentage of blockholders. Both regression models show that increased blockholder ownership is positively related with the probability of restatement. It is expected that blockholders demand high returns and because of their concentration of ownership, they would have more power than the average shareholder. The debt to assets ratio is also useful. With a positive coefficient of 0.1 and 0.11 in Model 1 and Model 2, respectively, it shows that with an increase in the debt to assets ratio, the probability of restatement increases. This is logical

because as debt increases, companies face increased pressure to produce positive financial performance to maintain credit and market standing. The pressure may cause the company to follow questionable accounting practices that could eventually lead to a restatement when policies are found to be incorrect. This is evidenced by positive coefficients of the debt to assets ratio. The results of assets and debt to equity ratio variables, coefficients of 0.02 and -0.01, respectively, produce a negligible impact on the results of both models, minimizing their importance.

### Conclusion:

Following the adoption of the Sarbanes-Oxley Act of 2002, the frequency of financial statement restatements has increased because of the increased scrutiny on company internal control practices. Previous literature has discussed the effects of restatements on both the company and investors. It was found that a company is likely to experience a decreased stock price while investors may suffer from a loss in value of their investment.

This paper seeks to find the economic determinants of restatements using a sample of 30 restatement firms derived from companies that have received SEC Enforcement Actions. These firms were matched by industry and asset size with 30 firms that had not restated financial statements. The sample was tested using four hypotheses covering the areas of debt to equity ratio, proportion of management ownership, proportion of blockholder ownership, and the use of a Big 5 accounting firm for auditing.

It was concluded that a negative relationship exists between audit by a Big 5 accounting firm and the incidence of restatement. Further, the use of Ernst & Young as an audit firm significantly decreased the incidence of restatement in this sample. Results for other hypotheses were not significant for this sample, possibly due to the small sample size.

The results were consistent with previous studies that found Big 5 firms, such as Ernst & Young, provided higher quality audits than non Big 5 firms. Also, although not significant, the negative coefficient of management ownership and the positive relationship of blockholder ownership also provide useful information. These groups both act in their own self interest – that of protecting the value of investments and earning higher returns, respectively. Further, the positive coefficient of the debt to assets ratio highlights the pressures that debt can cause for a company.

As Sarbanes Oxley provisions continue to be fully implemented, incidence of restatements will probably continue to remain higher than that of previous years. This will continue to affect both investors and the company itself, as restatements can be the catalyst for a further sequence of effects, such as stock price changes, credit rating revisions, and monetary fines. Research from this paper shows that the use of audit services from a Big 5 firm can mitigate these effects.

### Supplemental Tables:

Table 1:

Years Represented by Restatements	
1999	5
2000	11
2001	11
2002	2
2003	0
2004	1
N	30

Table 2:

Industry	
Utilities (22)	1
Construction (23)	1
Manufacturing (31-33)	9
Retail trade (44-45)	3
Information (51)	6
Finance and Insurance (52)	5
Real estate and rental and leasing (53)	2
Professional, scientific, and technical services (54)	3
N	30

Table 3:

Reason for Restatement	
Improperly recognized revenue	12
Improperly recorded expenses	7
Treatment of losses	4
Accounts receivable/payable	3
Improperly recorded gain on sale	2
Other	2
N	30

Table 4:

	Auditor					
	Restatement		Match co.		Total	
	co.	%	co.	%		%
PWC	8	26.7%	7	23.3%	15	25.0%
Arthur Anderson	6	20.0%	7	23.3%	13	21.7%
EY	4	13.3%	7	23.3%	11	18.3%
KPMG	5	16.7%	3	10.0%	8	13.3%
Deloitte	3	10.0%	4	13.3%	7	11.7%
Other	4	13.3%	2	6.7%	6	10.0%
N	30		30		60	

Source: Proxy statement for period of restatement

Table 5:

### Overall Mean of All Causes of Restatement

	Restatement	Matched Companies
Assets to equity ratio	4.46	7.86
Current debt to equity ratio	0.26	0.58
Long term debt to equity ratio	1.22	3.02
Total debt to equity ratio	1.48	3.61
% of management ownership	17.2%	19.5%
% of blockholder ownership	17.2%	14.6%



Table 6:

## Mean of Most Common Causes of Restatement

	Improperly Recognized Revenue		Improperly Recorded Expenses	
	Restatement	Match	Restatement	Match
Assets to equity ratio	2.24	2.02	7.09	3.25
Current debt to equity ratio	0.05	0.04	0.29	0.16
Long term debt to equity ratio	0.48	0.67	2.99	0.33
Total debt to equity ratio	0.53	0.70	3.29	0.49
% of management ownership	12.9%	16.3%	21.4%	41.0%
% of blockholder ownership	19.1%	4.5%	17.1%	23.1%

Table 7:

## Model 1

Variable	Mean	Median	Std Dev	High	Low
Log Assets	3	3	1.16	6	1
Debt/Assets	0.33	0.18	0.40	2.06	0.00
Debt/Equity	2.55	0.72	9.46	72.19	0.00
% management ownership	18.37%	9.77%	23.21%	100.00%	0.00%
% blockholder ownership	15.90%	10.85%	18.88%	99.50%	0.00%
Big 5 audit firm	1	1	0	1	0

## Model 2

Variable	Mean	Median	Std Dev	High	Low
Log Assets	3	3	1.07	5	1
Debt/Assets	0.33	0.18	0.40	2.06	0.00
Debt/Equity	2.55	0.72	9.46	72.19	0.00
% management ownership	18.37%	9.77%	23.21%	100.00%	0.00%
% blockholder ownership	15.90%	10.85%	18.88%	99.50%	0.00%
EY	0	0	0.39	1	0
Arthur Anderson	0	0	0.42	1	0
Deloitte	0	0	0.32	1	0
KPMG	0	0	0.34	1	0
PWC	0	0	0.44	1	0

Table 8:

## Regression Results

	Coefficient	P value
<b>Model 1</b>		
Intercept	0.74	0.03
Log Assets	0.02	0.75
Debt/Assets	0.1	0.57
Debt/Equity	-0.01	0.19
% management ownership	-0.16	0.61
% blockholder ownership	0.2	0.59
Big 5 audit firm	-0.36	0.17
Number of observations		60
<b>Model 2</b>		
Intercept	0.75	0.03
Log Assets	0.02	0.81
Debt/Assets	0.11	0.54
Debt/Equity	-0.01	0.17
% management ownership	-0.2	0.54
% blockholder ownership	0.22	0.56
EY	-0.5	0.11
Arthur Anderson	-0.39	0.19
Deloitte	-0.4	0.22
KPMG	-0.19	0.55
PWC	0.31	0.29
Number of observations		60

## Works Cited:

- Akhigbe, Aigbe, Kudla, Ronald J., & Madura, Jeff. (2005). Why are some corporate earnings restatement more damaging? *Applied Financial Economics*, 15, 327-336.
- Anderson, Kirsten L., & Yohn, Teri L. (2002). The effect of 10-K

restatements on firm value, information asymmetry, and investors' reliance on earnings. Working paper. Washington, DC: Georgetown University.

Bryan, Stephen, Lilien, Steven B., Ruland, William, & Sinnett, William M. (2005). Undoing the past implications of earnings restatements. *Financial Executive*, 21, 42-44.

Colbert, Gary & Murray, Dennis. (1999). State accountancy regulations, audit firm size, and audit quality: an empirical investigation. *Journal of Regulatory Economics*, 16, 267-285.

Hribar, Paul, Jenkins, Nicole T., & Wang, Juan. (2004). Institutional investors and accounting restatements. Working paper. St. Louis: Washington University

Kinney, William R., Jr., & McDaniel, Linda S. (1989). Characteristics of firms correcting previously reported quarterly earnings. *Journal of Accounting and Economics*, 11, 71-93.

Linn, Erik, & Diehl, Kori. (2005). Financial restatement causes, consequences, and corrections. *Strategic Finance*, 87, 34-39.

Moriarty, George B., & Livingston, Philip B. (2001). Quantitative measures of the quality of financial reporting. *Financial Executive*, 17, 53-56.

Palmrose, Zoe-Vonna, Richardson, Vernon J., & Scholz, Susan. (2004). Determinants of market reactions to restatement announcements. *Journal of Accounting and Economics*, 37, 59-89.

Richardson, Scott, Tuna, Irem, & Wu, Min. (2002). Predicting earnings management: the case of earnings restatements. Working paper. Philadelphia: University of Pennsylvania.

Wallace, Wanda A. (2005). Auditor changes and restatements. *The CPA Journal*, 75, 30-34.

## Faculty comments:

Dr. Don Finn explains the importance of his student's work in his letter of support for its publication. He said,

Ms. Berman's paper investigates the economic determinants of financial restatements that have been mandated by the Securities and Exchange Commission (SEC). Briefly, the SEC mandates that financial statements be restated when they find anomalies and/or errors (and possibly fraud) in the published information. The paper focused only on companies who have restated earnings and have been the target of an Enforcement Action. The purpose of identifying these companies is because they may have experienced fraudulent actions, rather than other less serious causes of restatements. To discern whether an accounting action is actually an economic determinant of restatement, all restatement firms were matched (a control group) with an equivalent non-restatement firm. The potential economic determinants between restatement and non-restatement companies were measured using the leverage, proportion of management ownership, and proportion of blockholder ownership.

It was determined that those firms audited by Big 5 firms were less likely to produce a restatement. In a second regression analysis, Ernst & Young was found to be the Big 5 firm whose clients were least likely to have a restatement. Thus, it was determined that large public accounting firms who performed audits were more reliable than those firms which are smaller and may not have the resources to properly audit large public companies.

## AN INVESTIGATION OF CHANGES IN CONTRIBUTIONS OF STATE LOTTERIES TO EDUCATION OVER TIME

By Andrea Lee Parker  
Department of Accounting

Advisor: Dr. Deborah Thomas  
Department of Accounting

### Abstract:

*Arkansas is one of many states in this country that is experiencing an education funding crisis. Despite the fact that states have started taking more responsibility for the funding of their public schools since the mid-1950s, litigation over education funding has occurred in almost every state in the United States. Litigation in Arkansas began in the 1980s and continues today with the Lake View case. Several alternatives have been proposed to reform the state's education system and its methods of funding, including school consolidation, raising taxes, and adopting an education-supporting lottery.*

*Lotteries have become very popular revenue raising mechanisms in the United States since the 1960s. Supporters of lottery adoption claim that lotteries are significant revenue raisers for new and existing programs and that they are better than increasing taxes. Opponents of lottery adoption contend that lotteries are regressive, implicit taxes and that the revenues from lotteries are fungible and do not increase education funding.*

*Data will be analyzed to determine whether lottery prize payout ratios have increased significantly over time, and whether changes in lottery prize payout ratios have had an impact on the money expended on public education. Based on this information and previous research, a recommendation will be made as to whether adopting an education-supporting lottery would be an effective and adequate way to help fund public education in the state of Arkansas.*

### Background:

#### *History of Arkansas Education Litigation*

The State of Arkansas' education system has been under fire since 1983, when *Du Pree v. Alma School District* was filed. In that case, the Arkansas Supreme Court ruled that there was "no rational relationship to education needs in the state's method of financing public schools." The Court also said that the variations in funding between the state's districts could not be justified by

the fact that the districts were "locally controlled." That ruling left the state with the ultimate responsibility to ensure an equitable public education system (*DuPree v. Alma School District*, 279 Ark. 340, 651 W.W.2d 90 (1983)).

Although the state did modify its funding formula in response to the Court's order, the education system was again tested in 1992 with *Tucker v. Lake View School Dist. No. 25*, and in 2000 with *Lake View School District v. Huckabee*. Lake View was the poorest school district in the state, and the district claimed that the way that the state allocated funds to school districts was inequitable. In 2001, a trial court in the state recognized that the education funding system was unconstitutional, saying that "The school funding system now in place . . . is inequitable and inadequate under . . . the Arkansas constitution." The court cited the Education Article of the state constitution, which declares that "the State has an absolute duty . . . to provide an adequate education to each school child." The court added, "Too many of our children are leaving school for a life of deprivation, burdening our culture with the corrosive effects of citizens who lack the education to contribute." The trial court ordered that, in order to be constitutional, the new funding system would have to be based on the amount of money needed to provide an adequate education for students (*Tucker v. Lake View School Dist. No. 25*, 323 Ark. 693, 917 S.W.2d 530 (1996); *Lake View School District v. Huckabee*, 351 Ark. 31, 78-79, 91 S.W.3d 472 (2002)).

According to the ruling of the Arkansas Supreme Court on November 21, 2002, which affirmed the trial court's ruling, an adequate educational system includes "adequately paid teachers, sufficient equipment to supplement instruction, and learning facilities that are [adequate]" (ACCESS Project-a, 2002). The State was given until January 1, 2004 to perform a cost study and create a constitutional funding system as outlined by the Arkansas Supreme Court (ACCESS Project-a, 2002). As a result of the state missing its January 2004 deadline to provide a constitutional funding system, Special Masters were appointed by the court. They maintain that "the Legislature [has] shortchanged

educational adequacy” and “the General Assembly [has] failed to make education its No. 1 priority” (Sadler, 2005).

#### *Costs of, and Proposals for, Reform*

The Arkansas Joint Committee on Educational Adequacy released a study in 2003 that estimated the additional amount of money that the state would have to spend to provide an adequate education to Arkansas’ students. The cost study found that annual funding would have to increase by \$848 million, to a total annual education budget of almost \$3.5 billion. At the time of the ruling, 41% of the State’s budget was dedicated to education (ACCESS Project-a, n.d.) In order to contribute an additional \$848 million per year to education, legislators and citizens alike have proposed many alternatives to raise additional revenue and/or cut costs in other areas. One of the most controversial solutions, backed by Governor Mike Huckabee, was a plan to consolidate schools that were under a minimum enrollment number. The Governor, using multiple studies on consolidation, proposed the plans because of the potential cost savings associated with merging small districts. The Governor’s plan was partially realized when the General Assembly passed legislation requiring the consolidation of districts with fewer than 350 students (ACCESS Project-a, n.d.). In addition to the consolidation plans, legislators suggested increasing sales taxes, which could bring in almost \$368 million in additional revenue per year. That amount would be almost half of the amount needed, as determined by the aforementioned cost study. Other proposals have included increasing other taxes, like income or property, and adopting a state-run lottery (Nelson, 2003; Robinson, 2003; Barnett, Ritter, & Lucas, 2003).

In October 2002, nearly a month before the Arkansas Supreme Court’s ruling in the *Lake View* case, researchers at the University of Arkansas conducted the fourth annual “Arkansas Poll.” Because of the immense attention and controversy surrounding the education crisis in the state, researchers used the poll to solicit Arkansans’ views concerning education reform and funding. In 2002, participants in the poll said that education was the second most important issue facing them. The nearly 800 Arkansans polled said that they would approve of such reforms as increasing teacher salaries, improving school facilities, and increasing vocational education opportunities. Arkansans were also asked about where they thought that the money needed for the reforms (at that time, it was estimated at over \$700 million) should come from (e.g. taxes, lottery adoption, cutting back government services.) Responses indicated that increasing taxes to fund education was the least desired option. Only two of the alternatives provided, adopting a state lottery and cutting back on government services, received a 50% or more approval rating. The researchers concluded that nearly 60% of Arkansans support the adoption of a lottery to help fund public education (Barnett, Summers, & Parry, 2003).

#### *Recent Arkansas Litigation*

In 2005, the Arkansas Supreme Court agreed to reopen the *Lake View* case after a motion was filed on behalf of dozens of Arkansas school districts. The motion claimed that the education budget approved by the legislature for the 2005-2006 school year violated the former *Lake View* opinion that required adequate funding for schools (ACCESS Project-a, n.d.).

Near the end of 2005, the Special Masters that had been appointed by the court submitted their recommendations in a “Findings of Fact” report. In summary, the Masters had several key points including: (1) education has not been the first priority of the state, (2) the amount of foundation funding for 2005-2006 should have been increased for a cost of living adjustment, (3) inefficiencies in the system could be best resolved through consolidation rather than by other approaches, (4) funds designated for school facilities are no where near the amount needed, and (5) the state is placing a larger responsibility of education funding on local governments and districts, and it is shrinking away from its own constitutional obligation (Bureau of Legislative Research, 2005).

A final decision from the Arkansas Supreme Court on the latest motion is still pending.

#### *Recent Arkansas Education Reform Proposals*

Democrat Bill Halter has recently announced that he will be running for Lieutenant Governor of the State of Arkansas. Since he announced his candidacy at the beginning of the year, Halter has laid out his personal education funding plan: adopting an education-supporting lottery. He has said that Arkansans are going to neighboring states to play the lottery anyway, and

“that the children of Arkansas should be the beneficiaries” (Blomeley, 2006). Halter claims that adopting a lottery could generate up to \$250 million per year for the state’s education system, a figure that is over four times the estimate given by the Arkansas Department of Finance and Administration’s (DFA) director Richard Weiss. The DFA’s estimate is based on lottery proceeds of states that are similar to Arkansas in size and composition (Thompson, 2006). Halter says that his proposal would send the money to college scholarships, pre-kindergarten programs, and teacher salaries. He also said that lottery revenue would add to, not replace, existing education funding (Sadler, 2006b).

In addition, Democrat Drew Pritt is currently working on a ballot proposal for the adoption of a lottery. His current proposal allocates most of the money to teacher salaries and school facilities construction. If the Attorney General approves the measure, and if Pritt can obtain enough signatures, the proposal may make it on the November 7<sup>th</sup> ballot (Sadler, 2006a).

#### *State Lottery Adoption and Success*

In 1964, New Hampshire introduced the first legal lottery of the 20<sup>th</sup> century. The measure was approved by 76% of voters

in a referendum (Hansen, 2005). Alicia Hansen, a staff writer for the Tax Foundation, says that the passage of the lottery was in part due to the fact that gambling was becoming more acceptable, while opposition to tax increases was growing. In addition, New Hampshire ranked the lowest in the nation in education spending, and the lottery was seen as a way to increase public school funding without increasing existing taxes or introducing new ones (e.g. a sales or income tax) (Hansen, 2004).

Today, forty-two states and the District of Columbia operate lotteries. In two of those states, Oklahoma and North Carolina, voters have just recently approved their operation. State-run lotteries are currently the most popular form of gambling in the nation, with more than half of Americans participating in recent years. In addition to being popular, state-run lotteries have also been very successful, with total spending on lotteries in the United States at nearly \$45 billion in 2003; with an average 31% takeout ratio (revenues less prizes), lotteries transferred about \$14 billion to state coffers that year (Hansen, 2004).

With over 80% of all states operating a lottery, and with the billions of dollars raised in revenue and the profits transferred to state coffers, state-run lotteries are extremely successful revenue-raisers. Currently, lotteries are used by some states to increase the money in their General Fund, while others have "earmarked" lottery profits for specific programs, such as economic development (Arizona and Oregon), the environment (Colorado and Nebraska), local food banks (Washington), and education. Education, by far, is the most popular program for which lottery profits are earmarked, with 24 of the 42 states earmarking profits for this purpose. Below is a map of the United States, with designations for the states that earmark proceeds for education (Hansen, 2004; Education Commission of the States-b, n.d.).

Many states, especially larger states and those that have been in operation for a longer period of time, have recorded astonishing transfers to state coffers and programs since their lotteries began. For example, New York, whose lottery began in 1967, and Texas, whose lottery began in 1992, have transferred over \$21 billion to education and \$9.7 billion to education and the General Fund, respectively, since their lotteries' inception. Many programs, including local food banks, education, and libraries, are being funded by and are benefiting from the adoption of a lottery (Hansen, 2004). Table 1 provides general information on each state's lottery, including how the lottery was approved and what programs are supported by lottery proceeds.

#### *Pros and Cons of the Lottery*

There are many economic benefits that can be realized from adopting a lottery to raise money for education and other worthwhile programs. Some of the arguments used in supporting lottery adoption are addressed below.

- The lottery is a "voluntary" tax
- The lottery provides funding for new and existing programs

- Lotteries are significant revenue raisers
- Lotteries keep citizens' money in-state

Despite the benefits provided by a lottery, lotteries are not without their opponents. There are many economic and social costs that may be incurred as a result of lottery adoption. Some of the arguments used in opposition to lottery adoption are addressed below.

- The lottery is not a voluntary tax
- Earmarking lottery money does not work because of the fungibility of state money
- Per pupil education spending may or may not increase
- The lottery is immoral
- The lottery may produce compulsive gamblers
- The lottery is a regressive implicit tax
- Lottery money is not a significant amount in state budgets
- Lottery revenues are unpredictable
- Lottery revenues may be unsustainable
- Lottery adoption may decrease public support for other education funding methods

#### **Method:**

The purpose of this research paper is to investigate two questions. The first question is to determine if state lottery prize payout ratios have changed significantly over time. The second question is whether lottery prize payout ratios significantly impact education spending in states with lotteries.

To answer the first question of whether prize payout ratios have increased over time, we have developed a regression model. The model is:

$$\text{LOTTPERPRIZE} = \text{YEAR} + \text{LOTTYEARS} + \text{NEIGHBOR} + \text{STATEPOP} + \text{PERCAPREV}$$

Where:

LOTTPERPRIZE is the state's prize payout ratio,

YEAR is the calendar year,

LOTTYEARS is the number of years that the state's lottery has been in operation,

NEIGHBOR indicates whether the state has a bordering lottery state,

STATEPOP is the state's population, and

PERCAPREV is the per capita revenue received by the state.

To answer the second question of whether prize payout ratios have had an effect on the amount of money spent on education, we have created the following model:

$$\text{SCHSPEN} = \text{PAYRATIO} + \text{NEIGHBOR} + \text{ENROLL} + \text{NUMSCHOOL} + \text{TAXBURDEN} + \text{GNP}$$

Where:

SCHSPEN is the education spending from all sources,

PAYRATIO is the prize payout ratio of the state's lottery,

ENROLL is the number of students enrolled in school within the state,

NUMSCHOOL is the number of school districts within the state,

TAXBURDEN is the per capita tax burden of the state's citizens, and

GNP is an economic indicator for the United States.

The data for the above variables (with the exception of the variables discussed below) were obtained from the Statistical Abstracts of the United States for the years 1977 to 2002.

The LOTTYEARS variable data was determined by obtaining the year of lottery adoption and adding a year for each additional year to determine the lottery's age. The variable data reported in a given year is the cumulative number of years of lottery operation for each state.

Data from the LOTTYEARS variable was used to determine the input data for the NEIGHBOR variable. A map of the United States was used to determine each year if the states that had lotteries were bordered by any other states that had lotteries. Data for the variable were coded with a "1" or a "0," with "1" meaning that there was a bordering lottery state and "0" if there was not.

The TAXBURDEN variable data were derived from using State Tax Collections information from the Statistical Abstracts of the United States and dividing by the STATEPOP variable data, which was also retrieved from the Statistical Abstracts.

For each regression model, three analyses were completed. The first is for all observations (i.e. all states), the second is for states that earmark lottery proceeds for education, and the third is for states that do not earmark lottery proceeds for education.

## Results:

The results for the first regression model, measuring whether lottery prize payout ratios have changed over time, are displayed in Table 1. The results for the second regression model, measuring whether education spending has changed over time, are displayed in Table 2.

## Lottery Prize Payout Ratios

The analysis of all states in the first model together revealed significant relationships between the dependent variable LOTTPERPRIZE and two independent variables, YEAR and LOTTYEARS. As expected, there is a significant positive relationship between time and prize payout ratios; as time passes, payout ratios tend to increase. Similarly, there is a significant positive relationship between the maturity of the lottery and payout ratios. That is, as lotteries age, payout ratios tend to increase. This result supports the idea that as lotteries mature, lottery commissions increase lottery prizes to keep interest and maintain revenues (Mikesell, 1987).

As for the analysis of states that earmark lottery proceeds for education, the YEAR and PERCAPREV independent variables were found to have significant relationships to LOTTPERPRIZE. PERCAPREV has a significant negative relationship with lottery prize payout ratios. As per capita state revenue increases, lottery prize payout ratios tend to decrease.

The analysis of states that do not earmark lottery proceeds for education revealed no significant relationships between the LOTTPERPRIZE variable and the independent variables.

## State Education Expenditures

The second regression model was designed to determine whether various independent variables affect the amount of money expended on public education. There were several significant relationships discovered in the second regression model when all states were analyzed. The results revealed a significant negative relationship between PAYRATIO and SCHSPEN. Three significant positive relationships between SCHSPEN and independent control variables ENROLL, TAXBURDEN, and GNP were also found.

The significant negative relationship between PAYRATIO and SCHSPEN indicate that as prize payout ratios (PAYRATIO) increase, the amount of money expended on education (SCHSPEN) decreases. This can be expected because although increases in prize payout ratios are thought to increase lottery revenues, the increased outflow of revenue to prizes may negatively affect the amount of money transferred to state coffers and education programs. An increase in the percentage of lottery income designated for costs (i.e. lottery prizes) would mean a lower percentage of income would be transferred to state coffers and education.

As the number of students enrolled in public schools increases (ENROLL), it is expected that states will have to increase their total education spending (SCHSPEN) to accommodate the additional students. This expectation is confirmed by the significant positive relationship seen between education expenditures and the number of students.

Similarly, it is expected that as each person in the population pays more taxes (TAXBURDEN) that a portion of that additional money will be used to support public education (SCHSPEN.) This is a reasonable expectation given that state education programs represent the single largest part of states' budgets. This expectation is confirmed by the significant positive relationship demonstrated between education expenditures and per capita tax burden.

It is expected that as GNP increases, state education spending (SCHSPEN) will also increase. As the production and output of the country or a specific state increases, it can be expected that the increased benefit realized from that income will be used (at least in part) to fund programs such as education. This expectation is confirmed by the significant positive relationship shown between education expenditures and GNP.

The results of the regression models for states that do and do not earmark lottery proceeds for education mirror the results for the group of all states as discussed above.

## Discussion:

The results of the current study should be considered as Arkansans determine whether or not a lottery would be an adequate and effective way of funding public schools within the state. Several significant and important relationships from the results that may impact citizens' decisions regarding lottery adoption will be discussed next.

### Lottery Prize Payout Ratios

The results that the passage of time and lottery age are significantly related to prize payout ratio were expected. As time passes and the lottery matures, lottery players will expect a higher return on the money that they spend playing lottery games. As players lose interest in lottery games where prize payout ratios remain stable over time, state lottery commissions increase these ratios to renew interest and to maintain and increase ticket sales. These results are consistent with the results of a study discussed earlier, "Evaluating the Life Cycles of Education-Supporting Lotteries." The results of the study indicate that while lottery revenues continue to increase, they are increasing

**TABLE 1**  
Results for Regression Equation Regarding  
Changes in Payout Ratios (LOTTPERPRIZE)

	<u>All States</u> Coefficient (t-statistic)	<u>Earmarking</u> Coefficient (t-statistic)	<u>Not</u> <u>Earmarking</u> Coefficient (t-statistic)
INTERCEPT	<b>-9.85552</b> (-4.43)	<b>-14.64738</b> (-6.59)	-3.18797 (-0.80)
YEAR	<b>0.00523</b> (4.65)	<b>0.00765</b> (6.80)	0.00186 (0.93)
LOTTYEARS	<b>0.00139</b> (2.60)	0.00066816 (1.45)	0.00234 (1.95)
NEIGHBOR	-0.00554 (-0.25)	-0.00590 (-0.30)	-0.02554 (-0.61)
STATEPOP	1.93203E-7 (.32)	-4.299E-7 (-0.89)	0.0000031 (1.32)
PERCAPREV	-0.000011 (-1.36)	<b>-0.0000167</b> (-2.07)	0.0000013 (0.08)
ADJUSTED R-SQ	0.0985	0.2593	0.0561

Variables with significant differences (measured at the .05 level) are denoted by their appearance in bold and *italics*

**TABLE 2**  
Results for Regression Equation  
Regarding Education Spending (SCHSPEN)

	<u>All States</u> Coefficient (t-statistic)	<u>Earmarking</u> Coefficient (t-statistic)	<u>No</u> <u>Earmarking</u> Coefficient (t-statistic)
INTERCEPT	1326338 (0.86)	<b>16981971</b> (3.37)	-840796 (-0.90)
PAYRATIO	<b>-10971405</b> (-3.92)	<b>-47474899</b> (-4.85)	<b>-3889122</b> (-2.29)
NEIGHBOR	-465765 (-0.63)	66974 (0.04)	-292016 (-0.62)
ENROLL	<b>5410.9172</b> (38.74)	<b>5978.395</b> (20.44)	<b>5366.90855</b> (28.57)
NUMSCHOOL	-401023 (-0.67)	<b>-5124523</b> (-3.34)	-196956 (-0.40)
TAXBURDEN	<b>1646.40675</b> (4.84)	<b>6378.74266</b> (5.62)	<b>1031.96396</b> (5.24)
GNP	<b>0.00029882</b> (4.01)	-0.00033759 (-1.47)	<b>0.00022502</b> (4.89)
ADJUSTED R-SQ	0.9027	0.9067	0.8736

Variables with significant differences (measured at the .05 level) are denoted by their appearance in bold and *italics*



at a diminishing rate. This effect is also stronger for older lotteries which are reaching maturity in their product life cycle. It can be reasonably concluded that one factor that is keeping lotteries growing is the increasing of prize payout ratios and other marketing techniques such as new product introductions. However, while there is no limit on the number of new products that may be released to encourage lottery players to participate, there may be a limit on the percentage of lottery revenues that can be reasonably paid out in prizes. Eventually, prize payout ratios will no longer be able to be increased and still maximize revenue, which may partly lead to the decline in real lottery sales as predicted by the study's authors (Mason, Steagall, Shapiro, & Fabritius, 2005).

Surprisingly, the presence of a neighboring state with a lottery did not significantly influence prize payout ratios. This was not the expected outcome, given Mikesell's findings that the presence of a neighboring lottery state and lottery maturity are the two factors that most influence lottery revenues. Due to his findings, it was expected that the competition of a neighboring lottery state would cause a state to increase its lottery's prize payout ratio in order to outperform neighboring competitors (Mikesell, 1987).

#### *State Education Expenditures*

The results of the second regression model have many implications for states that currently have lotteries supporting education and those states that are considering adopting an education-supporting lottery.

There has been a debate over whether lottery prize payout ratios should be increased or decreased to maximize revenue and proceeds to education. The results of the current study indicate that increasing prize payout ratios will decrease the amount of money expended on education. Thus, before a state decides that its lottery's prize payout ratios should be increased in order to maximize revenue, it should determine whether the increase in payout ratio will increase or decrease the lottery *proceeds* (or profit) to education. States should focus more on maximizing profits instead revenues from lotteries. Lotteries have been approved by voters because it is believed that the proceeds from the lottery will benefit important state programs such as education. Thus, increasing the amount of money that can be used to support programs (i.e. the lottery's profit) should be viewed as more important than increasing the amount of lottery revenue.

#### *Implications*

The analysis of the two regression models taken together leads to an interesting conclusion. First, the model measuring prize payout ratios indicates that payout ratios are increasing as a function of both time and lottery maturity (among other variables.) The passage of time and lottery age are two variables which cannot be controlled. Second, the model measuring education spending indicates that education spending is inversely

related to lottery prize payout ratios. That is, as lottery payout ratios increase, total education spending decreases, regardless of whether lottery proceeds are earmarked for education or used for some other purpose. The results of the two models suggest that lottery prize payout ratios will increase as a function of time and maturity, and these increases in prize payout ratios will tend to decrease total education spending over time.

The results of the study create many questions which should be considered before a lottery is adopted in the state of Arkansas. First, can Arkansas effectively compete with neighboring states if a lottery is adopted, and would its prize payout ratio be high enough to compete with neighboring states? Similarly, would the amount left over after prizes are paid be enough to help fund education? Currently, four of Arkansas' five bordering states have lotteries. Two of the states' lotteries, Missouri and Louisiana, have been in operation for 21 and 15 years, and have payout ratios of 65% and 55%, respectively. The other two neighboring lotteries, in Tennessee and Oklahoma, have only been recently adopted, and information from the states' lottery websites report current prize payout ratios of 58% and 52%, respectively. Thus, the state of Arkansas could reasonably expect to pay out between 50-60% in prizes in order to effectively compete. States that currently have payout ratios within that range transfer approximately 30 to 40% of lottery revenues to education or other programs.

Another question concerns how much revenue and proceeds to education could be expected from lottery adoption in the state of Arkansas. As discussed earlier, estimates of lottery profits from the proposed Arkansas lottery range between \$60 million (Arkansas Department of Finance) and \$250 million (Arkansas Lieutenant Governor candidate Bill Halter.) The large discrepancy between the estimates should be investigated, and an updated estimate of the amount of money that could be raised through a lottery should be made based on states with similar composition and demographic variables. Although the revenue and profit estimates provided by Bill Halter and the Arkansas Department of Finance and Administration are promising, the estimated increase in education spending may not be as much as expected. As discussed above, Arkansas' neighboring states that have lotteries and relatively high prize payout ratios may decrease the profits that the Arkansas lottery and education system would realize, as Arkansas would have to raise its prize payout ratio to remain competitive. Using the results of the current study, an increase in the prize payout ratio can be expected to negatively impact money expended on education.

In addition, will the amount of money raised through a lottery make a difference in education spending, when the "extra need" determined by the cost analysis is in the hundreds of millions of dollars? Given a reasonable estimate of potential lottery revenues and profits to education, Arkansans must determine whether lottery adoption is the best way to increase funding for public education. Although many may view the

revenues as "extra money" paid through a "voluntary tax," other research on lottery states' experiences should be considered. Research on lottery data previously discussed has found that lottery revenues only make up on average 2.25% of any state's general revenue. Still other evidence suggests that even earmarking lottery proceeds for education may not actually increase education spending; instead, lottery profits may replace the money currently spent on education.

If it is determined that lottery revenues will not make a significant difference in education spending, Arkansans and state legislators should question whether there is a better way to increase the amount of money available in the state's budget to spend on public education, such as tax increases or further consolidation of schools. While the current study did not intend to determine whether another method of funding would be better for increasing and maintaining education spending within the state, the results of this study imply that raising taxes may increase education funding. As formerly discussed, one of the control variables in the second regression model, tax burden per capita, had a significant positive relationship with total public school spending. Thus, while there are mixed results as to whether lottery revenues and profits would be sustainable and whether they would increase education spending over time, the results of the model imply that increasing taxes, although unpopular with the state's citizens, may be a more adequate and effective way of funding public education within the state.

### Conclusion:

A review of the literature has shown that the education funding crisis is not unique to the state of Arkansas, and it demonstrates that states have gone to different lengths in order to solve their funding problems. Research from the Arkansas Poll conducted in 2002 revealed that education is a top priority for Arkansans, and that the lottery is the most widely supported mechanism for increasing funding of public education in the state. While there are advantages and disadvantages to the lottery, both social and economic, forty-two states currently have lotteries, all of them reporting their lotteries as successful revenue-raisers. Arkansas can learn from the mistakes and successes of the states that have already adopted lotteries, those that are mature and the states that have only recently jumped on the lottery bandwagon. Arkansans should carefully weigh the pros and cons of an education-supporting lottery and should consider the economic and social costs and benefits associated with lottery adoption.

The results of the current study imply that over time as lottery prize payout ratios increase, education spending tends to decrease. While the results may be disappointing for those in favor of adopting an education-supporting lottery, the results point the state towards a funding method that may be more effective at increasing education funding: raising taxes. Both of these funding methods deserve further review by state legislators and citizens, especially as the Arkansas General Assembly begins its next legislative session in April 2006.

### References:

- ACCESS Project-a, Columbia University (n.d.). ACCESS Homepage. Retrieved March 17, 2006, from <http://www.schoolfunding.info/>
- ACCESS Project-b, Columbia University. (n.d.). *Arkansas Supreme Court declares school funding system unconstitutional; State must prepare adequacy study and "chart a new course for public education."* (2002). Retrieved from <http://www.schoolfunding.info/states/ar/11-21-02supremecourt.php3>
- Apuzzo, M. (2002). Critics pan lottery plan. *South Coast Today*. Retrieved March 28, 2006 from <http://www.s-t.com/daily/01-02/01-18-02/a01sr006.htm>
- Augenblick, J. (n.d.). School finance and property taxes: A continuing, if changing, linkage. National Center for Education Finance, National Conference of State Legislatures. Retrieved August 6, 2005, from [www.ncsl.org/programs/educ/NCEFTaxes.htm](http://www.ncsl.org/programs/educ/NCEFTaxes.htm)
- Barnett, J., Ritter, G., & Lucas, C. (2003). Educational reform in Arkansas: Making sense of the debate over school consolidation. Public Policy Working Paper Series, 2. Retrieved December 17, 2005, from [http://policy.uark.edu/paperseries/issue2\\_2002-03.html](http://policy.uark.edu/paperseries/issue2_2002-03.html)
- Barnett, J., Summers, J., & Parry, J. (2003). Analysis of the 2002 Arkansas Poll. *Office for Education Policy: Working Paper Series*, 1. Retrieved December 17, 2005, from [http://www.uark.edu/ua/oep/workingpapers/AR\\_Poll\\_02.pdf](http://www.uark.edu/ua/oep/workingpapers/AR_Poll_02.pdf)
- Blomeley, S. (2006, January 22). Governor hopeful outlines his goals Halter says he's for a state lottery. *Arkansas Democrat Gazette*. Retrieved March 10, 2006 from Lexis Nexis database.
- Bureau of Legislative Research. (2005). Summary of Special Masters Report. Retrieved March 21, 2006, from <http://170.94.58.9/data/education/MEMO-10-04-05LakeView-SummaryofMastersReport.pdf>
- Clotfelter, C. T., Cook, P. J., Edell, J. A., Moore, M. (1999). State lotteries at the turn of the century: Report to the National Gambling Impact Study Commission. Retrieved February 1, 2006, from [http://www.indiangaming.org/library/studies/1011-statelotrept\\_to\\_ngisc.pdf](http://www.indiangaming.org/library/studies/1011-statelotrept_to_ngisc.pdf)
- Cornwell, C. M., Mustard, D. B., Sridhar, D. J. (2001). The enrollment effects of merit based financial aid: Evidence from Georgia's HOPE Scholarship. *Journal of Labor Economics*, forthcoming. Retrieved February 20, 2006 from <http://www.terry.uga.edu/hope/research.html>
- Education Commission of the States-a (n.d.). ECS Education Policy Issue Site: Finance. Retrieved September 15, 2005, from <http://www.ecs.org/html/issue.asp?issueid=48>
- Education Commission of the States-b (n.d.). ECS Education Policy Issue Site: Finance—Lotteries. Retrieved September 15, 2005, from <http://www.ecs.org/html/issue.asp?issueid=48&subIssueID=47>
- Evans, W. N. & Zhang, P. (2005). The impact of earmarked lottery revenues on state educational expenditures. Retrieved December 17, 2005, from [http://www.bsos.umd.edu/econ/evans/wpapers/evans\\_zhang\\_earmark\\_lottery.pdf](http://www.bsos.umd.edu/econ/evans/wpapers/evans_zhang_earmark_lottery.pdf)
- Garrett, T. & Lawson, R. (1998). Earmarking tax revenues for education: Now you see them, now you don't. Retrieved September 15, 2005, from <http://www.buckeyeinstitute.org/article.php?id=292>
- Georgia Student Finance Commission. (n.d.). *Georgia's HOPE Scholarship and Grant Program*. Retrieved February 1, 2006, from <http://www.gsfc.org/hope/>
- Hansen, A. (2004). Tax Foundation Background Paper No. 46: Lotteries and State Fiscal Policy. Retrieved December 18, 2005, from <http://www.taxfoundation.org/news/show/65.html>
- Hansen, A. (2005). State-run lotteries as a form of taxation. Speech given at the National Coalition Against Legalized Gambling's 12<sup>th</sup> Annual Conference, October 8, 2005. Retrieved from February 7, 2006, <http://www.taxfoundation.org/news/show/1126.html>
- Hyden, D. S. (n.d.). The proposed Tennessee lottery: Pros and cons. Retrieved December 19, 2005, from <http://www.tba.org/lottery.html>
- Internal Revenue Service. (n.d.). *Internal Revenue Service Glossary*. Retrieved February 1, 2006, from [http://www.irs.gov/app/understandingTaxes/jsp/tools\\_glossary.jsp](http://www.irs.gov/app/understandingTaxes/jsp/tools_glossary.jsp)

Lottery Post. (2005, May 5). *New Florida lottery law increases prizes*. Retrieved December 17, 2005 from <http://www.lotterypost.com/news-113738.htm>

Lottery Post. (2003, September 10). *Paul recommends maximizing lottery prizes*. Retrieved December 17, 2005 from [www.lotterypost.com/threads73445.htm](http://www.lotterypost.com/threads73445.htm)

Mason, P., Steagall, J., Shapiro, S., & Fabritius, M. (2005). Evaluating the life cycles of education-supporting lotteries. *Public Finance Review*, 33, p 255-271.

Mikesell, J. L. (1987). The effect of maturity and competition on state lottery markets. *Journal of Policy Analysis and Management*, 6, p. 251.

Miller, D. & Pierce, P. (1997). Lotteries for education: Windfall or hoax. *State and Local Government Review*, 29, 34-42. Retrieved December 17, 2005, from <http://www.saintmarys.edu/~ppierce/pierce51.html>

National Center for Educational Statistics. (1999). *School finance litigation, by year, case, and status, by state: 1970-99*. Retrieved February 8, 2006, from <http://nces.ed.gov/edfin/litigation/>

Nelson, M. (2003, October 8). Legislators consider various plans. *Arkansas Democrat-Gazette*. Retrieved October 10, 2003, from Lexis-Nexis database.

Pulliam, C. (2003). North Carolina lottery for education: What are the odds our schools would win? Retrieved December 10, 2005, from <http://www.advocatesfored.org/publications/Lottery%20Paper%202003.pdf>

Robinson, D. (2003, October 10). Huckabee rolls out reform plan: Critics not convinced. *Fayetteville Morning News*, p. 1A.

Rychlack, R. J. (1992). Boston College Law Review: Lotteries, revenues, and social costs, A historical examination of state-sponsored gambling. Retrieved December 17, 2005 from <http://www2.tlc.ttu.edu/cochran/Gaming%20Law/rychlakbc3.htm>

Sadler, A. (2005, October 4). Masters: Legislature regressed on education funding. *The Morning News*. Retrieved October 20, 2005, from <http://www.arkansasnews.com/archive/2005/10/04/News/329239.html>

Sadler, A. (2006, January 11). AG rejects lottery, education proposals. *The Morning News*. Retrieved February 1, 2006, from <http://www.arkansasnews.com/archive/2006/01/11/News/332699.html>

Sadler, A. (2006, January 22). Halter launches campaign; proposes lottery. *The Morning News*. Retrieved February 1, 2006, from <http://www.arkansasnews.com/archive/2006/01/22/News/333231.html>

Skinner, R. A. (2005). State of the states. *Education Week Washington*, 24, p. 77-78, 80.

Staley, E. (2001). The public policy value of state lotteries. Retrieved January 13, 2005, from [http://www.rhodes.edu/public/20-Academic/2\\_1\\_5-politicalscien/pdfs/The%20Public%20Policy%20Value%20of%20State%20Lotteries\\_Frinn%20Staley.pdf](http://www.rhodes.edu/public/20-Academic/2_1_5-politicalscien/pdfs/The%20Public%20Policy%20Value%20of%20State%20Lotteries_Frinn%20Staley.pdf)

Thompson, D. (2006, January 24). Halter's lottery figures appear high compared to other states. Retrieved February 1, 2006, from <http://www.arkansasnews.com/archive/2006/01/24/News/333248.html>

Tierney, J. (2000). 'Money per pupil' is an incomplete response. *The New York Times*, p. B4. Retrieved February 1, 2006, from Lexis-Nexis database.

University of Arkansas Cooperative Extension Service. (n.d.). *A citizen's guide to Arkansas state taxes*. Retrieved February 7, 2006 from [http://www.arsba.org/Tax\\_Guide.pdf](http://www.arsba.org/Tax_Guide.pdf)

## Faculty comments:

Professor Deborah Thomas, Ms. Parker's mentor, explains the method and value of her student's research. She says,

Ms. Parker's interest in the potential value of a state lottery to support state programs began in a political science class, where she learned of the recent education funding litigation in our state. She approached me about investigating a tax aspect of this issue. After extensive background reading and discussions, we decided to investigate the use of lotteries for funding education.

This is a timely issue for the State of Arkansas. The Lakeview Supreme Court ruling and the recent determination of inadequate funding by court-appointed special masters have increased the pressure on the Arkansas state legislature to find additional support for the state's schools. Proposals for the adoption of a state lottery are re-emerging. Bill Halter, candidate for lieutenant governor, is one of many promoting a lottery to fund education ([billhalter.com/Issues#excellent](http://billhalter.com/Issues#excellent)). A petition to amend the state constitution to allow a state-run lottery and other types of gambling has been certified by the Secretary of State for the next general election ([www.sos.arkansas.gov/elections\\_006\\_amendments.html](http://www.sos.arkansas.gov/elections_006_amendments.html)).

Ms. Parker spent countless hours compiling the data for this project. She searched Statistical Abstracts of the United States for a twenty-five year period (1977-2002) to extract information about education and lotteries for each state. The resulting database includes hundreds of items of information that we hope can be used for future research.

Ms. Parker's initial research questions are: (1) Have lottery payout ratios increased over time? We hypothesize that as lotteries age, payout ratios increase in order to continue to attract players; and (2) Have changes in lottery payout ratios impacted the amount spent on education? We hypothesize that as payout ratios increase (decrease), the amount spent on education in the state will decrease (increase). We also investigate whether results differ between states that earmark lottery profits for education and those that do not.

Ms. Parker's statistical analysis indicates that lottery payout ratios have increased over time, particularly for states that earmark lottery proceeds for education. She also finds that payout ratios are negatively related to education spending in a state; as payout ratios increase, education spending in the state decreases. These primary findings, as well as other results, are of interest to state policy makers as they consider instituting a lottery as a source of funding for education.

## SEMI-STRONG FORM MARKET HYPOTHESIS: EVIDENCE FROM CNBC'S JIM CRAMER'S MAD MONEY STOCK RECOMMENDATIONS

By Elizabeth Dodson  
Department of Finance

Faculty Mentor: Dr. Craig Rennie  
Department of Finance

### Abstract:

*Mad Money has become one of the most popular shows on CNBC. The host, Jim Cramer, has an outlandish style and personality that viewers find intoxicating. Cramer's goal for the show is to make people money. Does he succeed? This paper finds that investors can expect to gain above-average, risk-adjusted returns by following Cramer's stock recommendations and trading accordingly. These findings challenge the semi-strong form market hypothesis. According to this hypothesis investors should not recognize gains trading on public information since it states that the market has already adjusted prices for that information. It also contributes to current literature by providing analysis on the different segments of the Mad Money program and serving as a jumping-off point for future research on a possible Jim-Cramer-Mad Money hedge fund strategy.*

### Introduction:

CNBC hit a homerun with its decision to put Jim Cramer on the air in all of his glory—hurling chairs across the floor, screaming at the camera, and ripping the heads off of bull-shaped stress balls, and that is in the first 5 minutes. A typical *Mad Money* show consists of Cramer starting off a pick of the day or week or describing recent market news and its possible impacts. Later in the show, Cramer opens the lines up for callers to run their stock picks by Cramer. During this “lightning round”, the pace is feverishly fast and “Booyah’s” (Cramer’s coined phrase for excitement and also a term of greeting) fill the air. Since its debut on March 14, 2005, Cramer’s *Mad Money* has become one of the top ranked shows on CNBC. The timeslot of 6 PM EST had previously been the lowest-rated slot until Cramer showed up with a vision. His vision was for a show with one simple goal, make people money. According to Derek Hoggett, editor and publisher of InvestmentWizard.com, “James Cramer is emerging as the most influential stock market guru since Dan Dorfman and Joe Granville.” Can Cramer really provide above-average, risk-adjusted returns for his viewers or is he dangerous to investors’ financial health?

This paper attempts to answer the question of whether Jim Cramer, utilizing his television venue, provides investors with a way to gain above-average, risk-adjusted returns. If Cramer is successful, then these results would challenge semi-strong form market efficiency theory. Semi-strong form market efficiency states that at any time stock market prices fully reflect all public information. Following this, investors should not be able to make better returns than the market trading on information that is already publicly available. Cramer does not provide investors with proprietary information. The support he gives for his choice of stocks comes from a variety of sources such as press releases, company websites, market news, and SEC filings. He is a major proponent of his viewers researching the companies he recommends before deciding to follow his advice. This paper contends, however, that regardless of following Cramer’s do-your-homework advice, investors can expect to see above-average returns simply by buying when and what he says to buy and selling when and what he says to sell.

This paper makes two contributions to current literature. First, it documents holding period returns that would have been achieved by following Jim Cramer’s investment strategy. Second, this study has implications for future hedge fund strategies. Perhaps a hedge fund could repeatedly beat its competition in the growing market for hedge funds by following Cramer’s investment advice and trading accordingly. This paper provides a jumping off point for other academics to continue the research and development of the probability and possible effects of a Jim-Cramer-Mad-Money trading strategy. As will be discussed in Section 2.2 of the Literature Review, one other paper has attempted to document the results of Cramer’s investment recommendations. This paper, however, takes a different slant than that of the Northwestern paper by investigating the different portions of the program (lightning round vs. non-lightning round) and hypothesizing that the two time periods are statistically significant from each other.

The remainder of this paper is organized as follows. Section II discusses current relevant literature and the development of

the hypotheses for the study. Section III describes the sample selection and the design for this study. Section IV lays out the results of the study. Section V provides the results of the sensitivity analysis. Section VI concludes.

## Literature Review and Hypothesis Description:

### 1. Efficient Market Theory

Semi-strong form market efficiency states that at any time stock market prices fully reflect all public information. If this is true, investors should not be able to obtain above-average returns trading on information after it is made public. "This hypothesis implies that investors who base their decisions on any important new information *after it is public* should not derive above-average risk-adjusted profits from their transactions" (Reilly & Brown, 2003). There is much disparity in the academic world as to whether the market truly is efficient at any level (strong, semi-strong, or weak) and is still up for debate. If Fama is correct and stock prices fully reflect publicly available information, then there is no opportunity for individual investors to try to find gains in trading securities. Regardless of the academic support for EMH, the actions of many investors reveal that they believe market inefficiencies do exist and that there are opportunities for abnormal gains. While many believe that investors are naïve to think they can actually beat the market because of its efficiency (Tam, 2002), individuals continue to tune in daily to Cramer's *Mad Money* looking for an advantage to do just that—beat the market.

The efficient market hypothesis was first developed by Eugene Fama in 1965. He states that "in an efficient market, competition among the many intelligent participants leads to a situation where... actual prices of individual securities already reflect the effects of information based ... on events that have already occurred" (1970). Fama's theory launched a new way of thinking about economic markets and led to much controversy. Perhaps the one puzzle in the discovery of whether the efficient market theory holds is whether investors really are intelligent as Fama surmises.

In his paper, Daniel (1999) asserts that "self-confident individuals will appear to be more competent than individuals who are insecure about their own abilities." Jim Cramer presents himself as a self-confident individual. It is no wonder that investors catch on to this intoxicating energy and trust in his interpretations of market information. Could Cramer be too confident? Daniel supports his argument saying that overconfidence can have both a direct and indirect impact on how individuals process information. If investors adopt his overconfident attitude, they could overweight his suggestions and neglect to search their own base of knowledge to make the most intelligent investment decision. However, if they jump on the bandwagon too quickly they are not accurately following Cramer's advice. Cramer advises his viewership to "do their homework" and research a company before making the decision

to buy the stock. This paper contends that regardless of following his do-your-homework advice, investors can gain positive returns by following his I-want-you-to-buy-this-stock advice—perhaps investors are more intelligent than originally surmised if they are already trading with Cramer's stock recommendations.

### 2. Results of Other Event Studies

Event studies are used to examine abnormal returns surrounding a particular economic event and provide a test for the EMH. Popular event study topics for testing semi-strong market efficiency theory include stock splits, initial public offerings (IPOs), exchange listings, unexpected world events and economic news, announcements of accounting changes, and corporate events. According to Reilly and Brown (2003) the evidence from tests of the semi-strong EMH draws mixed conclusions. Numerous event studies on a range of topics like stock splits, exchange listings, and initial public offerings provide support while numerous studies on predicting the expected return over time or for a cross section of stocks actually provide evidence count to semi-strong efficiency.

However, there are multiple event studies that offer evidence to counter semi-strong market efficiency. One such study is that of Engelberg, Sasseville, and Williams (2006). They find that Cramer's recommendations do have an effect on stock prices in the short run concluding that Cramer's show caused mispricing in the market. Their findings document the existence of inefficiencies in the market—negating the semi-strong form market efficiency hypothesis. However, this study's findings differ from those of Engelberg, et al. because this study differentiates between the different portions of the television programs. This study also uses TheStreet.com as the primary data source, but Engelberg et al. used a different second-hand source other than the Personal Finance Blog discussed later in the paper.

Two additional studies that tested the value of investment advice looked at the recommendations made through the *Heard on the Streets (HOTS)* column in the *Wall Street Journal*. The first study was conducted by Davies and Canes (1978). They study the recommendations presented in the *HOTS* articles and found that the dissemination of information from a primary source (the analysts) to a public format can significantly affect stock prices. As a follow up to this study, Liu, Smith, and Syed (1990) extend the study using a more recent sample. They also find that investment advice has economic value due to the observed abnormal returns for both buy and sell recommendations on the day of articles publication.

### 3. Cramer's Reviews

After Cramer graduated from Harvard in 1977, he began work as a journalist at the *Los Angeles Herald Examiner*. He returned to Harvard, received his Juris Doctor degree, and took a position with Goldman Sachs' Sales and Trading department.

He began his own hedge fund company (Cramer Berkowitz) with his partner Jeff Berkowitz in 1987. Though his hedge fund was hugely successful, Cramer left in 2000 citing anger-management and stress issues as the dominant factor to his decision. Shortly after leaving Cramer Berkowitz, he began appearing as a host on *America Now* and *Kudlow & Cramer*. After *Kudlow & Cramer*, Cramer began his own show which mixed his reputation for an anger-management problem and his stellar stock-picking record into a dynamite television program that has soared to the top of the charts.

Cramer's *Mad Money* came under scrutiny on June 19, 2005, when the New York Post printed an article by Richard Wilner entitled "Smart 'Money'? Cramer's Bark Worse Than Bite." The article explained that while Cramer's show is entertaining, it is not the best investment advice. Cramer responded to this article by offering his own, "Cramer's Mad Money Record Speaks for Itself." In his article, Cramer provides the first week's worth of his recommendations and their results. He wraps up the article by saying "I thought you would enjoy the actuals, though, and you might conclude that my bite and my bark are both pretty effective." The first week's worth of recommendations did look impressive, but that was not enough to keep those who doubt Cramer's success at bay.

On February 27, 2006 TheStreet.com revealed that it and Cramer had received a subpoena from the SEC in response to stock manipulation allocations made by Overstock.com's CEO Peter Byrnes (Byrnes denied he was behind the SEC probe). The accusation is that a group of hedge fund managers and journalists conspired to drive down the stock price in order to gain from the stock's fall in price. Cramer responded by saying he was the target of the investigation because "I said the stock was going lower. I didn't get the subpoena because I'm corrupt, I got it because I tried to get people out of a stock that we said was going lower, and went lower." He subsequently wrote bull on the subpoena and dropped it on the floor.

### *Hypotheses Development*

Based on the analysis of current literature and other relevant event studies, two hypotheses for this test were developed. The first hypothesis is as follows.

**H<sub>0</sub>:** investors can not gain above-average, risk-adjusted returns for the 5 and 26 day trading windows by following the recommendations made by Jim Cramer during his CNBC show *Mad Money*

**H<sub>1</sub>:** investors can gain above-average, risk-adjusted returns for the 5 and 26 day trading windows by watching *Mad Money* and following the stock recommendations

The second hypothesis was made by further dividing the data and distinguishing between when the recommendation was made during the program—either lightning round or non-lightning round. The hypothesis for this division was developed as follows.

**H<sub>0</sub>:** regardless of when the recommendation was made during the program, investors can not expect to gain above-average, risk-adjusted returns for the 5 and 26 day trading windows

**H<sub>1</sub>:** the two different segments of the program can gain the investor differing levels of above-average, risk-adjusted returns for the 5 and 26 day trading windows when one considers whether the recommendation was a buy or sell

### *Sample selection and study design:*

The data for this test was collected from two internet data sources. The first and most reliable is TheStreet.com, a website co-founded by Jim Cramer. This website provides recaps of the daily shows from the most recent date back to June 28, 2005. For data earlier than June 28, I relied on the Personal Finance Blog (PFBlog.com) where a faithful viewer reviews in detail Cramer's daily recommendations. The variables I considered were the date of each show, company name, Cramer's position on the stock, and whether he was recommending the stock during the normal show time or the adrenaline-pumped lightning round. The data collection required some subjective decision making. In instances when it was difficult to distinguish, I relied on my ability to read his comments and decide whether he was bullish or bearish on the stock. The rule of thumb used was if he spoke about the stock in a positive light, it received a "buy" recommendation—alternatively, negative spins on a stock received a "sell" recommendation.

The next step was to define the window of time. Since the show has only been on the air for a little over a year, the analysis would have to be short-term. This being the case, the decision was made to focus the analysis on 26-day trading window (five days prior to the event and twenty days after). Since *Mad Money* is a daily show, Cramer will often times repeat his recommendation for a stock within the created constraint of 25 days. Therefore, the process then became to remove "doubled-up" recommendations that could skew the results. Repeated recommendations were deleted unless Cramer changed his views on the stock, up to 25 days after the initial recommendation. After the 26-day window, the same recommendation on a stock was considered a unique event and included in the analysis. After controlling for repetition, there were 3,550 observations. The time period extended from April 19 to November 30, 2005. The data could not be extended beyond its end date since 20 days after the trade recommendation were required for the analysis. Also, data on stock prices and indices were only available through 2005.



After the time period of the study was established, the second set of data needed to be collected. This portion of the study included finding the stock returns for each company in Cramer's basket of recommendations as well as the returns for the value-weighted index, the proxy for the market's return. Through the University of Arkansas's subscription to Wharton Research Data Services (WRDS) the returns were accessed for the value-weighted index using datasets from the Center for Research in Security Prices (CRSP). The value-weighted daily returns include all distributions, on a value-weighted market portfolio (excluding American Depository Receipts (ADRs)). All returns were found for the 26-day trading window.

Since the alternative hypotheses of this paper are that investors can expect to earn above-average risk-adjusted returns, the daily returns of each stock had to be adjusted for the returns an investor could expect to make from investing in a portfolio of the market during the same time period. The cumulative abnormal returns were calculated for each of the 26 days surrounding the event as:

$$\text{Cumulative Abnormal Return} = \text{Portfolio Return}_t - \text{Return on Value-Weighted Index}_t \text{ (where } t=0 \text{ is the recommendation event)}$$

The data underwent two tests of significance for the two time periods evaluated for the existence of short-term abnormal gains one could make from watching *Mad Money*.

The first test looked at the entire 26-day trading window (5 days before the event, the day of, and 20 days after). The second time period was for the five days surrounding the recommendation event (two days before, the day of, and two days after). Tables 1 and 2 summarize the observations for the two time periods and the two tests for each hypothesis (buy and sell only and lightning round vs. non-lightning round). From the study, one can conclude that Jim Cramer's stock recommendations made during his television program, *Mad Money*, can gain investors cumulative abnormal returns.

## Results:

The conclusion that investors can gain above-average, risk-adjusted returns is statistically significant at a 95% level of confidence. The variables were tested twice for the given hypotheses. The variables for the first test were all buy and sell recommendations. These variables were significantly different from each other with a p-value of less than 0.0001 for both time periods. The second set of tests determined the significant difference of lightning round buy and sell recommendations and non-lightning round buy and sell recommendations. For the 5-day period variables were found to be significantly different from each other with p-values less than 0.0001 for all variables. The 26-day time period found statistically significant differences for all variables with the following p-values: 0.0042 for lightning round buy vs. lightning round sell and 0.0002 for non-lightning

round buy vs. non-lightning round sell. The final test compared lightning round and non-lightning round recommendations. These tests were significantly different at a 95% level of confidence only for the 5-day time period with p-values of less than 0.0001 for buys and 0.0059 for sells. Table 3 summarizes the significant difference for the variables in each time period test.

Investors can earn abnormal returns from following Cramer's trading advice. They can expect to gain the highest returns by following his recommendations for a position on a stock during the non-lightning round portion of his show. For the 26-window trading period, this class of recommendations averaged a 2.4% cumulative abnormal return as opposed to 1.4% for non-lightning round recommendations. Sell recommendations made during the regular part of the show averaged -1.2%—for non-lightning round sell recommendations, the CAR was just over 0.25%. The largest CAR (3.6%) occurred two days after the recommendation event for buys on stocks recommended during the non-lightning round portion. The second largest was the day before, or one day after the announcement. Following the second day, returns slowly begin to decrease dropping by a total of 1% by the twenty-fifth trading day in the study for non-lightning round buy recommendations. Figure 1 illustrates the variables and cumulative abnormal returns.

Another phenomenon witnessed within these tests was that the CAR for the recommendations increases initially before the day of recommendation. There are many possible reasons why this is the case. One is that Cramer bases his recommendations off of the market news—information that is publicly available—and the market is adjusting for this public information. Another possibility is that there is a leak in information. This conclusion concerns the idea that Cramer's recommendations actually affect the stock price which is a consideration that should be looked into in the future. The leak, therefore, means that investors trade with the expectation of being able to ride the change in stock price due to Cramer's recommendation. Regardless of the reason, investors are still able to gain returns thanks to Cramer's recommendations.

## Sensitivity Analysis/Robustness Check:

For robustness, equally-weighted daily returns, including all distributions on an equally-weighted market portfolio (including ADRs), were also collected from CRSP. The cumulative abnormal returns were calculated for each of the 26 days surrounding the event as:

$$\text{Cumulative Abnormal Return} = \text{Portfolio Return}_t - \text{Return on Equally-Weighted Index}_t \text{ (where } t=0 \text{ is the recommendation event)}$$

Conclusions drawn from new tests using this index instead of the Value-weighted index were similar to those based on the value-weighted index.

## Conclusion:

The energy of Cramer's show is invigorating. While his enthusiasm has developed many strong supporters and believers, he has also managed to draw plenty of criticism for his personal style and investment advice. He maintains, though, that the objective of the show is to make people money. As long as he remains loyal to the show's original purpose, he is doing his job to the best of his ability. The results of this study support just that. There are some shortcomings of this study, though that could potentially diminish the returns an investor can expect to make. These factors include considerations like transaction costs and time concerns. Taking these factors into consideration could absolve any abnormal returns observed in the study.

Even with the above considerations, the findings that investors can expect to gain above-average, risk-adjusted returns by following Cramer's recommendations on *Mad Money* make a contribution to the research on the semi-strong form market efficiency hypothesis. If this hypothesis holds firm, investors can not expect to gain cumulative abnormal returns trading on information after it is made public—the information is already priced into the stock. However, this paper contends that not only can investors find gains in an inefficient market but they can gain these returns by following the recommendations made by Jim Cramer during his show *Mad Money*. Other studies, such as that conducted by a group of academics at Northwestern, also find that Cramer is able to provide investors with gains over those of the market. The findings of this paper, coupled with that of Northwestern, provide additional support that the market is not operating at a semi-strong form of market efficiency. These findings make contributions to the current literature by providing information about the returns gained from the different portions of the show. There are also implications for future research into

the development of a hedge fund strategy based off of this assessment.

## Author's note:

I wish to thank the following people for their dedication and support throughout this long and sometimes stressful process. Thanks for making my last semester a memorable and worthwhile experience: Dr. Craig Rennie, John Norwood, Jeff Jones, Candice Norton, Rivka Berman, Gary Schubert, Leslie Lemiso, and Susannah Rodgers.

## References:

- Cramer, J. J. (2005, June 20). Cramer's mad money record speaks for itself. *The Street.com*. Retrieved January 5, 2006 from [http://www.thestreet.com/funds/smarter\\_up/10228758.html](http://www.thestreet.com/funds/smarter_up/10228758.html)
- Daniel, K. & Titman, S. (1999). Market efficiency in an irrational world. *Financial Analysts Journal*, 55, 28-43.
- Engelberg, J., Sasseville, C., & Williams, J. (2006). Is the market mad? evidence from Mad Money. Northwestern University: working paper.
- Fama, E. (1965). Random walks in stock market prices. *Financial Analysts Journal*, 51, 75-81.
- Fama, E. (1970). Efficient capital markets: a review of theory and empirical work. *The Journal of Finance*, 25, 383-417.
- Graham, J. & Campbell, R.H. (1996). Market timing and ability and volatility implied in investment newsletters' asset allocation recommendations. *Journal of Financial Economics*, 42, 397-421.
- Liu, P., Smith S.D. & Syed, A.A. (1990). Stock price reactions to the Wall Street Journal's securities recommendations. *The Journal of Financial and Quantitative Analysis*, 25, 399-410.
- Lloyd Davies, P. & Canes, M. (1978). Stock prices and the publication of second-hand information. *The Journal of Business*, 51, 43-56.
- Tam, F. (2002, November 4). Markets not entirely efficient. *Business Times*, 7.
- Reilly, F.K & Brown, K.C. (2003). Investment analysis and portfolio management (7th ed.). Ohio: Thomson.
- Wilner, R. (2005, June 19). Smart 'money'? Cramer's bark worse than bite. *New York Post*, 32.

Table 1: Description of Data for 5-day and 26-day Trading Window Buy and Sell Recommendations Only Value-Weighted Index:

This table provides the descriptive statistics for the entire data set. The two variables considered in this test were the cumulative abnormal returns for all buy and sell recommendations made by Cramer during *Mad Money* television broadcasts from April 19, 2005 to November 30, 2005. Stock recommendations that were repeated within the 25-day limit are not included in the sample. The two time periods tested—the 5-day and 26-day trading windows—are represented in the following table.

Recommendation	Variable	N	Mean	Median	Std Dev	Minimum	Maximum
Sell	CAR for 5 days	1661	-0.0029	-0.0023	0.0663	-0.5864	0.5713
	CAR for 26 days		0.0045	0.0043	0.1381	-0.8182	1.0170
Buy	CAR for 5 days	1889	0.0155	0.0073	0.0625	-0.1945	0.8581
	CAR for 26 days		0.0214	0.0152	0.1092	-0.5453	0.8467

Table 2: Description of Data for 5-day and 26-day Trading Window Buy/Sell Recommendations and Lightning Round vs. Non-Lightning Round Value-Weighted Index

This table provides the descriptive statistics for the entire data set. The four variables considered in this test were the cumulative abnormal returns for buy and sell recommendations made during lightning round and non-lightning segments for *Mad Money* broadcasts from April 19, 2005 to November 30, 2005. Stock recommendations that were repeated within the 25-day limit are not included in the sample. The two time periods tested—the 5-day and 26-day trading windows—are represented in the following table.

Recommendation	Variable	N	Mean	Median	Std Dev	Minimum	Maximum
Sell/Non-lightning Round	CAR for 5 days	143	-0.0170	-0.0041	0.1034	-0.5864	0.4065
	CAR for 26 days		-0.0174	-0.0093	0.1670	-0.8062	0.5735
Sell/Lightning Round	CAR for 5 days	1518	-0.0016	-0.0021	0.0616	-0.3081	0.5713
	CAR for 26 days		0.0065	0.0050	0.1350	-0.818	1.0170
Buy/Non-lightning Round	CAR for 5 days	473	0.0335	0.0168	0.0830	-0.1945	0.8581
	CAR for 26 days		0.0268	0.0206	0.1201	-0.5453	0.7738
Buy/Lightning Round	CAR for 5 days	1416	0.0094	0.0040	0.0526	-0.1860	0.6370
	CAR for 26 days		0.0196	0.0127	0.1052	-0.4348	0.8467

Table 3: Test of Significant Difference for Cumulative Abnormal Returns Value-Weighted Index

This table provides a summary of the test of significance for the variables under the two tests of the hypotheses. Panel A refers to the test for the first hypothesis (whether investors can gain cumulative abnormal returns by following the recommendations made by Jim Cramer during *Mad Money*). The remaining panels offer evidence for the second hypothesis (the returns investors can expect to gain are dependent upon during which segment of *Mad Money* the recommendation is made). The two time periods tested—the 5-day and 26-day trading windows—are represented in the following table.

## Panel A: All Buy and All Sell Recommendations

Variable	N	Buy	Sell	Difference	p-value
CAR for 5 days	3550	0.0157	-0.0029	0.0184	<.0001
CAR for 26 days	3550	0.0214	0.0045	0.0169	<.0001

## Panel B: Lightning Round Buy and Sell Recommendations

Variable	N	Buy	Sell	Difference	p-value
CAR for 5 days	2934	0.0094	-0.0016	0.0110	<.0001
CAR for 26 days	2934	0.0196	0.0065	0.0131	0.0042

## Panel C: Non-lightning Round Buy and Sell Recommendations

Variable	N	Buy	Sell	Difference	p-value
CAR for 5 days	616	0.0335	-0.0170	0.0504	<.0001
CAR for 26 days	616	0.0268	-0.0174	0.0441	0.0002

## Panel D:

## Lightning Round Buy vs. Non-lightning Round Buy Recommendations

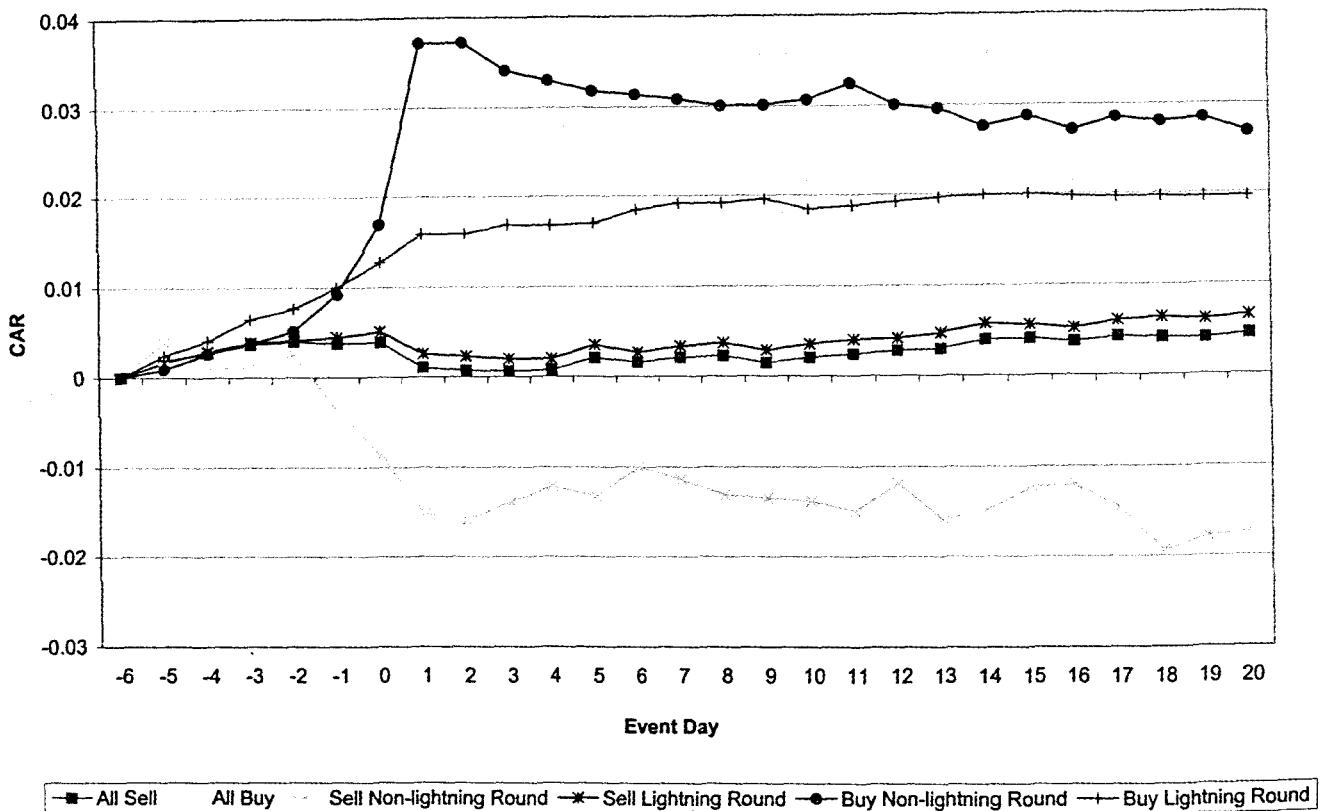
Variable	N	Lightning Round	Non-lightning Round	Difference	p-value
CAR for 5 days	1889	0.0094	0.0335	-0.0240	<.0001
CAR for 26 days	1889	0.0196	0.0268	-0.0072	0.2730

## Panel E: Lightning Round Sell vs. Non-lightning Round Sell Recommendations

Variable	N	Lightning Round	Non-lightning Round	Difference	p-value
CAR for 5 days	1661	-0.0016	-0.0170	0.0154	0.0059
CAR for 26 days	1661	0.0065	-0.0174	0.0239	0.0271

Figure 1: Cumulative Abnormal Returns for 5-day and 26-day Trading Window Value-Weighted Index

This figure provides a visual interpretation of the data analyzed in the study. The time period analyzed is five days before the recommendation event and twenty days after. The recommendation event day is time 0. The six variables represented in this figure are the two for the first hypothesis (whether investors can gain abnormal returns) which are all buy and sell recommendations made and the four variables for the second hypothesis (the amount of gain investors can expect to gain depends on when the recommendation was given during the show) which are buy and sell recommendations made during lightning round and non-lightning round portions



#### Mentor comments:

Dr. Craig Rennie made the following remarks about Ms. Dodson's research:

As Elizabeth Dodson's thesis advisor, I strongly recommend Elizabeth's research on Jim Kramer's *Mad Money* recommendations for publication in *Inquiry*. Her paper makes a material contribution to the empirical investments finance literature dealing with efficient markets and asset pricing by investigating the effectiveness of investment recommendations made by the immensely successful CNBC investments commentator of a hit TV show.

Traditionally, there has been debate in the academic investments finance literature about semi-strong form market efficiency. Prior literature generally shows investments recommendations from newsletters and other commentators generate marginally superior returns, but not necessarily after transaction costs are taken into account.

Elizabeth's research extends this work in a new way by investigating the effectiveness of recommendations made by a widely followed CNBC commentator who is a highly experienced former hedge fund manager. Only one other working paper exists on this topic, but it fails to differentiate between Jim Kramer's own recommendations and his commentary in the "lightning round" where he responds to questions from call-ins. Elizabeth notes that Jim Kramer's recommendations could predict stock price movements, or cause stock price movements. What matters to investors is that his recommendations can be used to achieve abnormal stock returns.

Importantly, Elizabeth's findings of almost a 3 percent abnormal return over a short term holding period are striking. Her research can be used by investors to realize abnormal stock returns. Her results also cast further doubt on the semi-strong form efficient markets hypothesis.

## BIOMETRIC SECURITY: ARE INEXPENSIVE BIOMETRIC DEVICES RELIABLE ENOUGH TO GAIN WIDE-SPREAD SECURITY USAGE?

By Brian Thanh Tran  
Department of Information Systems

Advisor: Dr. David E. Douglas  
Department of Information Systems

### Abstract:

*The ever growing need for security in today's world requires exploring the feasibility of various security methods to ensure the safety of the world's population. With the tremendous growth of technology, e-commerce, and business globalization, society implements new methods to try to battle security problems. Technology advances has resulted in a number of inexpensive biometric devices to the marketplace. Two questions surface regarding this devices—are they reliable enough for general usage and will people be willing to use them?*

*This research conducted a repeated design experiment to determine the effectiveness of four inexpensive biometric devices—three fingerprint readers and an iris scanner. Further, a questionnaire was designed to gain insights to the views of subjects using these biometric devices. On average, all the devices performed well for identification purposes—the fingerprint readers performing better than the iris scanner. The questionnaire revealed that most people prefer fingerprint readers over that of iris scanners and that although 60% of the people surveyed had heard of biometrics, only 21% of those surveyed had ever used a biometric device. The public does not feel that these devices provide complete security, but does provide a reliable means for identification.*

### Introduction

Biometrics is the study of biological characteristics and behaviors for the purpose of verifying identity. With the tremendous growth of technology to try to battle security problems, the reoccurring question often becomes “is this a reliable security method?”

Methods used by forensic teams such as latent fingerprints, DNA, hair samples, or fiber analyses are not considered to be in the field of biometrics. Biometrics has a key advantage over traditional methods such as tokens (smartcards, keycards, etc) and passwords because they are measurable and use physiological and/or behavioral characteristics to verify the identity of an individual. Tokens can be lost, stolen, forgotten, and in some

cases be duplicated. Passwords have the problems of being stolen, broken, shared, or forgotten.

With the continuation of corporate globalization, events such as 9/11 and the London bombings of July 2005, and identity theft, the need for better security measures have become more prominent and necessary. Recently, biometric technologies are becoming security options in everyday use for businesses and organizations. Trying to take a leap into the biometrics market, Accenture was given a \$10 billion contract in 2004 to incorporate biometric identification measures for the U. S. Visitor and Immigration Status Indicator Technology program, which allows for the tracking of foreigners entering the United States. With the growing importance of ecommerce and online transaction processing the security of IT infrastructure has never been as critical as it is now.

In the midst of the technology age, we are trying to find more methods in which to solve the problems of identity theft and verification to allow for a safer society. This project will provide reliable insights into using inexpensive biometric devices for identity and authentication. In analyzing the collected data, the wealth of information derived from the primary research will allow for a better understanding of how effective current biometric technology really is and what impact it could possible make in the present and future.

New technologies open a world of opportunities. Having an accurate identification and authentication process will help deter crimes, fraud, and save critical resources that can be used to advance the efficiency of society. Currently, the United States has about \$1 billion dollars in welfare benefits that are claimed by individuals who are double dipping with fake identities. Companies such as Mastercard estimate their credit card fraud to be approximately \$450 million annually, and ATM cards have a fraud worth of approximately \$3 billion annually ({Jain, A. 1999; 2}). According to Erik Bowman from CardTech/ SecurTech, the growing demand for network security industry will increase the market for biometric applications from \$24 million in 1997 to \$60 in 1999 ({Lawton, George 1998; 17}).

These facts demonstrate that the opportunities for biometric devices to enter the market and make a direct impact are very high.

However, the possibility for so many variables such as age, ethnicity, different body states such as sickness or allergies, or even medical solutions such as contact lenses and Lasik eye surgery, challenges using biometrics to verify identity in everyday life. The technology, if in fact successful, should provide improvements for identity theft and fraud problems. If properly designed, biometric devices could allow for technological advances to improve efficiency and productivity of society as a whole. Nevertheless, this technology also raises a number of questions, some of which are listed below.

Who uses biometrics?

How does biometrics work?

How does biometrics NOT work?

How effective are these technologies?

How will these technologies affect private lives?

### Purpose of Study:

The research will attempt to answer these questions by use of biometric products that can be purchased by the everyday consumer. Multiple devices testing the same biometric variable were purchased to confirm the reliability of the device and the variable they are testing. Biometric devices considered for the research project included those manufactured by companies such as BioCert, Microsoft, Panasonic, and APC. Through surveys, we hope to gain insight on how people view these devices and do a comparison on whether or not people find these devices as privacy's enemy or privacy's friend. This study and the survey focus on the inexpensive biometric devices and do not incorporate all biometrics.

The first research question addresses the accuracy of inexpensive biometric devices. The research question is answered via the three hypotheses in the next section. A questionnaire was developed to help answer the second research question.

To answer the first research question required taking multiple measurements on the same subject with each biometric device. In this research, each subject was identified at six different times with each of the four biometric devices. Thus, it is a repeated measures design model. In addition, age and gender were used as factors. The repeated design research model is shown below as Diagram 1.

Diagram 1. Diagram of Research Design.

ID #	Age	Sex		Run 1	Run 2	Run 3	Run 4	Run 5	Run 6

### The Repeated Measures Design and Hypothesis:

The repeated measures design applies when the values of the dependent variable(s) represent repeated measures on the same subject. The repeated measures are taken at different times on the same subject. This model is used extensively in medical research but also applies to this research.

The model allows testing of hypotheses about measurement factors – referred to as within-subject factors and includes interactions of within-subject factors with independent variables often called between-subject factors. In this model the between-subject effects represent the different biometric devices and the within-subject effects are the effects over time. The model also includes interactions.

The following hypotheses are used to test the major research question relating to inexpensive biometric devices:

**H<sub>1</sub>:** Fingerprint readers and iris scanners accurately identify people on the first try.

**H<sub>2</sub>:** The effectiveness of fingerprint readers and iris scanners do not change over time.

**H<sub>3</sub>:** Age and gender have no impact on fingerprint readers and iris scanner accuracy.

### Fingerprint Recognition Devices:

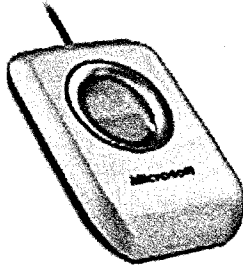
Fingerprints have long been known to be unique to every person. This being so, many places have used fingerprints as a way to identify individuals. As security concerns continue to grow, so does the number of passwords. Personal computers today often store sensitive and confidential data. They are also the access point to corporate networks. As systems become smaller and more mobile, they are more at risk of being lost or stolen. Biometrics provides users a convenient and secure way to manage and access multiple security phrases and codes.

Today's fingerprint recognition and identification systems work by taking a digital scan of a person's fingertips and then records the finger's unique physical characteristics. The fingerprint data will either be stored as an image or encoded as a character string, depending on the developer. The advancement of fingerprint identification has made it the technology of choice in today's consumer products, such as computer keyboards, cell phones, door locks and employee time clocks. Relative to other biometric choices, fingerprint recognition is cheaper, faster and accurate enough for most applications in which it is used. To prevent fooling the system, newer fingerprint identification systems also measure blood flow to the finger, so that a fake finger can't be used. Listed and shown below are the 3 relatively inexpensive fingerprint recognition devices that were used in this study.

Microsoft Fingerprint Reader – Model 1033.

Retail Price of \$50.

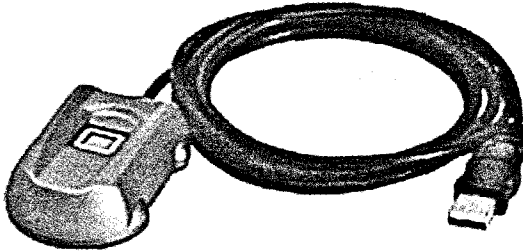
Average Attempt of 1.08 over 6 runs.



APC Biometric Fingerprint Reader– Model BioPodMP4.

Retail Price of \$50.

Average Attempt of 1.10 over 6 runs.



BioCert Fingerprint Reader – Model Hamster III.

Retail Price of \$130.

Average Attempt of 1.07 over 6 runs.



## Data Collection and Results for Fingerprint Recognition Devices:

### Microsoft Fingerprint Reader

From the data collected, we saw on average that there were no runs that require more than 2 scans to identify the individual. We also saw that adults 40 and up had a lower average on attempts for the device to recognize them versus that of adults 18 to 39. Observations were made that the reason this is probably

true is that the older group took more time when placing their finger upon the device. We also found out that there was not enough evidence to distinguish a difference in the attempts for males versus that of females.

### APC Fingerprint Reader

From the data collected, we saw on average that there were no runs that require more than 2 scans to identify the individual. The APC device however, recognized the adults 18 to 39 better than the adults 40 and up. We also found out that this device required less attempts to recognize the males than the females.

### BioCert

From the data collected, we also saw that on average there were no runs requiring more than 2 scans to identify the individual. The device manufactured by BioCert also required less attempts to recognize adults 18 to 39 versus that of the adults 40 and up. When comparing males to females, this device had the same average attempts for both sexes.

## Recommendations for Fingerprint Recognition Devices:

The purpose of this study was to test and compare the reliability of the current low cost biometric fingerprint recognition devices. As we can see, the results will vary according to brand. On average alone, the BioCert reader had the best at 1.07, closely followed by the Microsoft at 1.08. The APC ended up with an average of 1.10. However, the majority of the tested preferred the Microsoft reader, then the BioCert, and finally the APC reader.

This study has shown that on average, the current low cost providers of fingerprint recognition devices that were used in this study are reliable enough in which it can recognize an individual on average of less than 2 attempts. Common problems that were noticed that caused a person not to be recognized were fingers that were wet, had substances such as dirt or food on them, or cuts would result in a rejection. Also, residue that was left from the last person that used the device sometimes caused the device not to recognize an individual.

Possible users for this type of device are large corporations that have user and password log-ons that maybe stolen. Or the same corporation could implement such devices for a time clock to help prevent time clock fraud in which another person can punch another person's employee number in. Another possible use would be credit card companies having a fingerprint scan on the magnetic strip and when it is used, instead of a signature, fingerprint verification would be needed.



### Iris Recognition Devices:

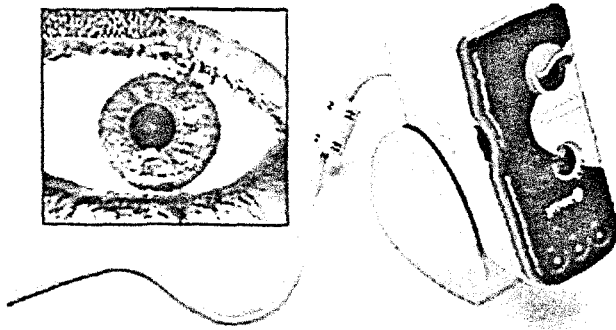
Iris recognition technology examines the unique features of the human iris, the colored portion of the eye, to create an image of the iris. This is then translated into a data template, which can later be used to identify individuals or authenticate user privileges. The iris of the eye possesses physical patterns unique to each person. Similar to fingerprints, no two irises are alike in the world. Iris recognition biometric systems can analyze over 200 points of the iris, including rings, furrows and freckles. Eyeglasses, contact lenses, and eye surgery do not change the characteristics of the iris.

This method of identification is becoming widespread, and is only second behind using fingerprints for identification due to its relative cost and accuracy. To prevent fooling the system iris recognition systems often vary the light in order to see that the pupil dilates, so that a fake eye can't be used. Due to continual advances and range of costs in biometric iris technology, not all of the devices that were originally planned to be used in this study were acquired. Shown below is the only iris recognition device used in this study because all the other iris recognition devices were too expensive.

Panasonic Authenticam – Model BMET100US.

Retail Price of \$200.

Average Attempt of 1.255 over 6 runs



### Data Collection and Results for Iris Recognition Devices:

The iris recognition device in this study also was able to recognize an individual on average in 2 attempts. However, adults 40 and up did have more problems than the adults 18 to 39 when it came to the device identifying the individual. When comparing males to females, both sexes had a very similar average in attempts required for identification.

### Recommendations for Iris Recognition Devices:

The purpose of this study was to test the reliability of the current low cost biometric iris recognition devices. As we saw in this test, iris recognition was not as reliable as fingerprint recognition. Also, this device was the least favorite among all the biometric devices that were used in the study. The Panasonic

Authenticam was often much slower in recognizing the individual and not as easy to use as described by most surveyed. While we had an average of less than 2 attempts for recognition, there is still plenty of room for development of a cost effective iris recognition device. Faster recognition and ease of use are among the top two.

Although eyeglasses and contact lenses do not change the characteristics of the iris, we did notice that they did affect the results. Eyeglasses tended to cause glare when the device tried to read the iris and sometimes were required to be taken off to get a good read on the iris. Currently, I would not recommend the use of a low cost iris recognition device as many people find them bothersome and are not very likely to accept them at their current state.

### Other Biometric Devices Not Used In Study:

Retinal scanning systems look at the pattern of blood vessels at the back of the eye. Retina scans use a light to shine on the retina, and require that the person place their eye close to the scanner, remain still, and focus on a specified location. Biometric retina recognition systems are among the most accurate of all biometric technologies and as such are used at military installations and other high-risk facilities. It is also quite expensive due to the hardware needed. Retica systems is currently the only full-eye biometric technology company.

Biometric facial recognition measures and analyzes the physical attributes of a person's face. Characteristics measured include the overall structure and shape of the face, and distances between the eyes, nose, mouth, and jaw edges. Face recognition systems can accurately verify the identity of a person standing two feet away in less than five seconds.

Biometric hand geometry recognition measures and analyzes the physical attributes of a person's hand. Characteristics measured include the overall size and shape of the hand, including the lengths of the fingers and joints, and characteristics of the skin such as creases and ridges.

Hand recognition systems are fairly common, however they are expensive due to the proprietary hardware and not that accurate compared to other technologies.

Speech recognition is another biometric technology that distinguishes an individual. The device is not the most accurate as a person's voice can change as different symptoms such as sickness or allergies appear. Speech recognition technology has been in development for a while, as right now it is commonly used to dictate text into the computer or to give commands to the computer (such as opening application programs, pulling down menus, or saving work). While the accuracy of speech recognition has improved over the past few years some users still experience problems with accuracy either because of the way they speak or the nature of their voice.

## Emerging Biometric Technologies:

Newer biometric technologies using diverse physiological and behavioral characteristics are in various stages of development. The biometric devices describe in this area are currently being developed and may emerge over the next 2 to 4 years, while others are many years from implementation currently only available commercially. There are a few that are available, but very limited and only to those who are willing to put the capital into further developing and research these devices. Each biometric method's performance, as with all biometric devices, can vary widely, depending on how it is used and its environment in which it is used.

One emerging biometric technology is facial thermography which detects heat patterns created by the branching of blood vessels and emitted from the skin. The patterns, known as thermograms, create a very unique image. Even identical twins have different thermograms. Developed in the mid-1990s, thermography works much like facial recognition, except that an infrared camera is used to capture the images. Currently the efforts into furthering this technology are on pause due to the high cost.

Researchers are investigating a biometric technology that can distinguish and measure body odor. This technology would use an odor-sensing instrument, an electronic "nose", to capture odor that is emitted through the skin's pores all over the body, which in return would make up a person's smell. However, distinguishing one individual's odor versus that of another may one day be a realistic, using this technology is currently very complex due to different variables that may take place such as the use of deodorants or perfumes. Different diets and medications can also influence the body odor emitted from a person and makes the development of this technology slow.

A popular route in today's market is combining multiple biometric measures into one device to ensure validity when taking a reading. Retica Systems, the only full-eye biometric technology company, is currently developing a handheld device that will compare both the retinal and iris to ensure that the individual being scanned is who they say they are.

Another technology currently in development is a vein scanning biometric technology that can automatically identify a person from the patterns of the blood vessels in the back of the hand. The technology uses near-infrared light to detect vein vessel patterns. Vein patterns are distinctive between twins and even between a person's left and right hand. Developed before birth, they are highly stable and robust, changing throughout one's life only in overall size. The technology is not intrusive, and works even if the hand is not clean.

The key distinction for biometric devices is a unique trait that can be measure. The exact composition of all the skin elements is distinctive to each person and makes it a prime

candidate for being a biometric measure. Skin has layers that differ in thickness and pigmentation differences that make each individual's measure unique. Skin pattern recognition technology measures the characteristic spectrum of an individual's skin. Current skin pattern recognition technologies use a light sensor to illuminate a small patch of skin with a near-infrared light. The light is then analyzed by a spectroscope and then a distinct optical pattern can be formed.

## Privacy Issues:

With any new security measure, the issues of personal privacy and invasion arise. This is especially the case when it comes to biometric devices and their uses for identification. People fear that their biometric readings will link them to their personal data or allow them to be tracked, in a "Big Brother" type situation. A common question that appears when the topic of biometrics comes up in reference to privacy is if the government or some other group or person could get a hold of their personal information if they had access to the biometric system. The common misconception is that the readings that are taken during the enrollment phase do not actually hold any personal information, but it is the relationship between the image and the database that holds and relates the personal information. More common identification methods such as driver's license reveal much more information than a biometric measure, and are much easier to steal or counterfeit.

## Survey and Data Results:

Listed in this section is a summary of the issues that were asked in the survey. The total numbered surveyed was 82 subjects. Subjects can be classified in these sets:

1. 53 subjects were between the ages of 18 and 39
2. 29 subjects were ages 40 and up.

---

1. 55 subjects were male.
2. 27 subjects were female.

---

1. 21 subjects were female and between the ages of 18 and 39.
2. 6 subjects were female and ages 40 and up.
3. 32 subjects were male and between the ages of 18 and 39.
4. 23 subjects were male and ages 40 and up.

The questionnaire revealed that most people prefer fingerprint readers over that of iris scanners and that although 60% of the people surveyed had heard of biometrics, only 21% of those surveyed had ever used a biometric device. The public does not feel that these devices provide complete security, but does provide a reliable means for identification.

When the issue of identity theft was asked in the survey, 67% of the women 40 and up and 61% of the men 40 and up felt that their identity was not secure. In comparison, 42% of the women 18 to 39 and 36% of the men 18 to 39 felt that their identity was not secure.

We also found that 53% of the women 18 to 39 and 45% of the men 18 to 39 view that the security of biometrics were secure. In comparison, 50% of the women 40 and up and 52% of the men 40 and up felt that biometric devices were secure.

When asked how likely they were to accept biometrics in day to day life, 57% of females 18 to 39 and 30% of the males 18 to 39 we willing to accept the usage of biometric devices. However, only 30% of females ages 40 and up and 48% of males 40 and up are willing to accept the usage of biometric devices into daily life.

### Conclusion:

Though currently not widely accepted as a reliable and secure method of identification, biometric devices have made great advances in both reliability and price. This study provides very valuable insights into using inexpensive biometric devices for identity authentication and how the public view these devices. This research concludes that the fingerprint devices are preferred over the iris recognition device and the fingerprint devices were more reliable than the iris scanner that was used in this study. We can also see that people are also favorable to the idea of using biometric devices to verify their identity when making credit card purchases. The majority of the people surveyed were also likely to accept biometric usage into daily life. Using the data collected from the device testing, we ran a statistical analysis through SAS and found that Time and Device are statistically significant but the interaction between the two is not.

Hypothesis 1: Was confirmed without statistical analysis. The average number of times required to identify a person averaged less than 1.5 for all devices—which rounds to 1. Thus, it can be concluded that the biometric devices accurately identify people on average on the first attempt. Hypothesis 2: Using SAS 9.1 and an alpha of .05, we discovered that time was significant factor. Hypothesis 3: Using SAS 9.1 and an alpha of .05, there is a significant difference between the devices and age. Further, the interaction of gender and device was also significant.

The possibilities for biometrics being implemented into society are limitless and only a few issues were addressed in this study. One of the main issues that will always come up when it comes to personal identification will be privacy, and the general public will fear that their information will be obtainable by all. This study that determined the public does not feel that these devices provide complete security, but does provide a reliable means for identification. Although biometrics does not completely solve the problems of identity theft, fraud, and security, it is a good step in trying to deter those problems.

### References:

- Jain, R. Bolle, and S. Pankanti, Eds., *Biometrics: Personal Identification in Networked Society*. Norwell, MA: Kluwer, 1999.
- Lawton, George. "Biometrics: A New Era in Security." *Computer* August 1998: 16-18.
- Lawton, George (1998, August). Biometrics: A new era in security. *Computer*, 31(8), 16-18. Retrieved February 1, 2006, from Research Library database.
- Souter, Colin. "Biometric System Performance and Security" 18 Oct. 2005 <[http://www.bioscrypt.com/assets/bio\\_paper.pdf](http://www.bioscrypt.com/assets/bio_paper.pdf)>
- Wayman J.L., Ed., "National Biometric Test Center Collected Works", August 2000. 18 Oct. 2005 <<http://www.engr.sjsu.edu/biometrics/nbtccw.pdf>>

### Faculty comments:

Dr. David Douglas said of his student's work,

Brian Tran conducted research on the science of biometrics for identification and security purposes. His research provides examples of business costs of incorrect identity and security. Certainly, security has been a focal issue over the past few years.

After a broad background study on the field of biometrics, Brian focused on the availability of inexpensive biometric devices. Advances in technology have spawned a number of inexpensive devices. Two questions surface regarding these devices—are they reliable enough for general usage and will people be willing to use them?

Brian designed a repeated measures experiment, with corresponding hypotheses, to answer the first question. This experiment used three fingerprint readers and one iris reader—all inexpensive devices. Age and gender were also factors in the experiment. His well designed experiment provides a basis for evaluation of the devices in terms of accurate identification for security purposes. This part of the research is important because not only were the inexpensive devices found to be accurate but the design provides a basis for further research with additional factors such as race and with other devices as they appear in the marketplace.

Further, Brian developed a questionnaire to capture the subjects' acceptance of such devices for security reasons. This information is valuable in two ways—used a gauge of whether the devices will be acceptable at this point in time and as a reference point to determine if attitudes change over time.

## **SECTION III: SCIENCE AND ENGINEERING**

**BIOLOGICAL AND AGRICULTURAL ENGINEERING,  
CHEMICAL ENGINEERING, COMPUTER SYSTEMS AND  
COMPUTER ENGINEERING, AND INDUSTRIAL  
ENGINEERING**

## EVALUATION OF RICE HULLS AS A LIGNOCELLULOSIC SUBSTITUTE IN WOOD PLASTIC COMPOSITES

P. Jack Bourne  
Department of Biological & Agricultural  
Engineering

Faculty Mentor: Dr. Sreekala Bajwa  
Department of Biological & Agricultural  
Engineering

### Abstract:

*Rising raw material costs and shortage of woody materials necessitate alternative sources for lignocellulosic material in wood plastic composites (WPC). This study was conducted to evaluate rice hull (RH), an agricultural residue, as a cellulosic substitute in WPC. Samples were fabricated with approximately 4% zinc stearate, 48% high-density polyethylene obtained from recycled plastics and 48% lignocellulosic material by mass. The composition of the lignocellulosic material was changed from 0 to 100% RH at 20% increments while the remainder was wood flour. The extruded samples were tested for mechanical properties such as specific gravity, water absorption, linear coefficients of thermal expansion, and strengths under compression, shear and bending. The results showed that increasing the proportions of RH to wood flour in the new composite increased the specific gravity but decreased the water absorption. The rice hull rates did not change any of the strength properties. Overall, physical and mechanical properties of the new composite was comparable to that of two of the commercial WPCs. Therefore, rice hull is a viable and renewable alternative for lignocellulosic material in WPC intended for non-structural applications such as decking, fencing, flooring and OEM.*

### Introduction:

Wood plastic composites (WPC) have the advantages of wood like appearance combined with low maintenance and durability for nonstructural applications. Wood flour is used as a lignocellulosic filler in a polymer matrix to form WPC. With diminishing forest resources, pressure from environmental groups and outsourcing of furniture manufacturing, the availability of wood flour has been diminishing. Accordingly, the price of wood flour has been steadily increasing, by as much as 300-500% in the past 5 years. The current reported price of wood flour is approximately 22-27 cents per kg. Demand of WPC has been steadily increasing in the last decade, with a market demand of over 1 billion kg in 2005 (Freedonia Group Inc. 2004). Considering

a typical composition of 50% wood in WPC, the current demand of 0.5 billion kg of wood flour is expected to grow to approximately 0.8 billion kg in 2009 based on the estimated demand of WPC at 1.7 billion kg (Morton et al. 2003). The price of filler (plastic) has been soaring in the past couple of years due to the steadily escalating oil prices. The growth in WPC demand and Although the mounting raw material prices have lowered the profitability margin significantly, the market demand for WPC has been steadily increasing. Therefore, it is imperative to find alternate sources of raw materials including the lignocellulosic filler and polymer in WPC. Agriculture waste is an untapped resource that can provide an inexpensive alternative for lignocellulose in WPC.

Rice hulls (RH) are a renewable agricultural waste that contains natural fibers. Global hull production has been estimated at 100 million metric tons annually, and the material is readily and cheaply available (Oliver, 2004). Rice hulls have been used as a fuel source for residential heating and as an energy source for industrial processes such as rice parboiling operations or electricity generation (Primenergy, 2006; The Agrilectric Companies, 2005). Rice hull is commonly included in animal bedding and used as a fiber source in feed products. Additionally, rice hulls have also been used in sustainable architecture as a loose-fill insulation having low cost and good flammability and insulative characteristics (Oliver 2004). Rice hull ash is suitable for commercial applications within the steel and cement industries. The ash has also been used in more limited quantities for applications as a soil amendment, pressing aid, insulating material, or oil absorbent (Bronzeoak, 2003).

Rice hulls have previously been used for the commercial production of structural plastic composites. Nexwood, a commercially available WPC used approximately 9.1 million kilograms of rice hulls for their WPC production in 2002 alone (Winandy 2004). However, this is not a prominent trend in the WPC industry. There is very little information available in literature regarding the incorporation of rice hulls in plastic

composites, and their impact on the physical and mechanical properties of the resulting composites.

Rice hulls form approximately 20% of the dried paddy on stalk. It is a good source of lignocellulose, with lignocellulose constituting approximately 60-78% of the weight (Luh, 1991). Bulk density of rice hull varies from 0.10 to 0.16 g/ml, and true density ranges from 0.67 to 0.74 g/cm<sup>3</sup>.

The high silica content of the hulls contributes to an effective hardness of 5.5 to 6.5 on the Moh's scale and yields abrasive properties (Luh 1991). As a heat source, rice hulls yield approximately 13.9 MJ/kg which is roughly one-third of that delivered by fuel oil (Cramer et. al. 1991). Chemical constituents of the rice hull are listed in table 1.

**Table 1: Chemical Composition of Rice hulls (Juliano, 1985; Miller and Eisenhauer, 1982; National Academy of Sciences, 1971 as cited by Luh, 1991)**

Component	Content
Moisture (%)	7.6-10.2
Crude Protein (%)	1.9-3.7
Crude fat (%)	0.3-0.8
Crude Fiber (%)	35.0-46.0
Available Carbohydrate (%)	26.5-29.8
Ash (%)	13.2-21.0
Silica (%)	18.8-22.3
Calcium (mg/g)	0.6-1.3
Phosphorous (mg/g)	0.3-0.7
Neutral Detergent Fiber (%)	66-74
Acid Detergent Fiber (%)	58-62
Lignin (%)	9-20
Cellulose (%)	28-36
Pentosans (%)	21-22
Hemicelluloses (%)	12
Total Digestible Nutrients (%)	9.4

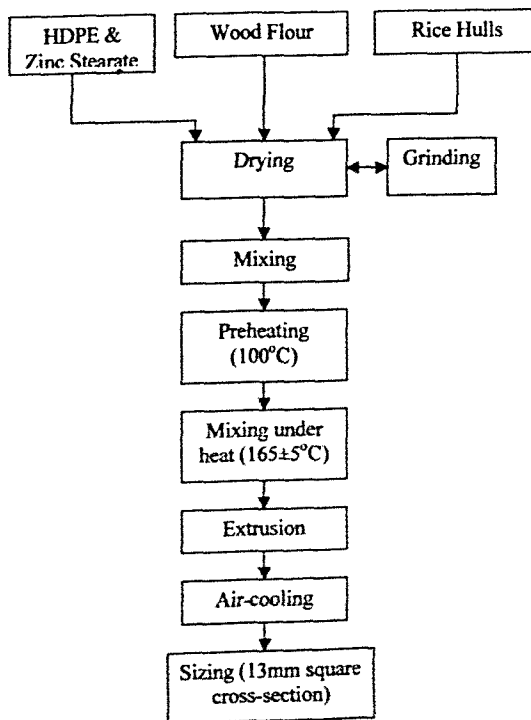
Arkansas is a major rice producing state, which produced approximately 50% of the total US production of 9.4 metric tons of rice each year. Rice hull is a byproduct produced during rice milling, which is available in plenty in Arkansas. Two of the major WPC manufacturers are also located in Arkansas. Therefore, alternate uses of rice hull will have value to the rice producers and processors in Arkansas. WPC manufacturers in Arkansas will benefit from using rice hull as a raw material in WPC production. Therefore, this study was undertaken with the goal of testing rice hulls as a substitute for wood flour in WPC. The specific objectives were to:

- (1) Develop a process for preparing lignocellulosic-plastic composite from rice hulls.
- (2) Evaluate the mechanical properties of this new wood-hull-plastic composite, and compare it with commercially available products.

## Materials and Methods:

### Sample Preparation

WPC samples were extruded from a mixture of high-density polyethylene (HDPE), raw or parboiled rice hulls (RH), wood flour and zinc stearate. Recycled plastic pellets were used as source of HDPE. Wood flour used in the study was made from shavings of southern yellow pine ground to a 40-mesh size. Zinc stearate was used as a lubricant. Rice hulls were obtained from Ricelands Inc. in Stuttgart. The experimental design consisted of 6 combinations of lignocellulose, 4% ZnSt and 48% HDPE. The six ratios of wood flour and RH were approximately 0:100, 20:80, 40:60, 60:40, 80:20 and 100:0 by weight. Due to resource constraints, varying numbers of samples were produced for each composition. Additionally, some samples were produced with only raw rice hulls while others contained only parboiled rice hulls. Statistical analysis indicated no significant difference in mechanical or physical properties of lignocellulosic composite samples made with raw and parboiled rice hulls (results not shown). Therefore, samples made with raw rice hulls and parboiled rice hulls were not separated in this study. The number and composition of samples under each treatment is shown in Table 2. For each composition, an adequate number of production runs were made such that the number of samples from separate batches used for each test corresponded to entries in the "Samples per Test" column of Table 2.



**Figure 1: Process Flow Chart**

Table 2: Recipe (compositions) of the lignocellulosic plastic composite.

Treatment	Samples per Test	Ingredients of the lignocellulosic plastic composite (g)			
		ZnSt	HDPE	Wood Flour	Rice Hull
Wood	3	13.3	160	160	0
RH10	5	13.3	160	128	32
RH20	2	13.3	160	96	64
RH30	2	13.3	160	64	96
RH40	2	13.3	160	32	128
RH50	4	13.3	160	0	160

A process was developed to fabricate lignocellulosic plastic composite samples from rice hulls, wood flour and plastic (Figure 1). Rice hulls and 40-mesh wood flour were dried separately in a convection dryer at 60°C until all evaporable water was lost. Rice hulls were then ground until passing through a 2 mm screen, at which time they were returned to the dryer. Lignocellulosic materials were stored in the convection dryer until use. For sample preparation, all ingredients were mixed in the correct proportion and preheated to approximately 100°C. The pre-heated mix was then manually mixed under heat until the material temperature reached 160±5°C. The samples were then extruded into blanks with a square cross-section of 19 mm using an arbor press. The extruded samples were air-cooled. To ensure equivalent surface qualities between batches, air cooled samples were planed on all sides to a final cross section of 13 mm square. Samples were then cut to the desired lengths for testing according to the recommendation of the ASTM standard followed.

The mixing, heating and extrusion chambers were fabricated in-house. The extruder was built from a hollow steel cylinder with an internal diameter of 62.4 mm and length of 103.5 mm. An end cap with the die was welded to one end of the cylinder. The die was fabricated from a seam welded square tube with a length of 23.7 mm and hole size of 19 mm. The inside surface of the die was polished to a fine finish to obtain a clean extrusion surface. Several trial runs were made before extruding the actual samples to ensure smooth operation of the die. The die was cleaned and polished after each batch of samples was extruded.

#### Sample Testing

The extruded samples were tested for specific gravity, water absorption, linear coefficient of thermal expansion (LCTE), bending properties, shear strength, and parallel compressive strength. Specific gravity of the samples was tested according to the ASTM D 6111-97 standard test method for bulk density and specific gravity of plastic lumber and shapes by displacement. Specific gravity tests were conducted as specified in the standard with the exception that measurements were made at 19°C. Water absorption by the samples was tested according to the standard

test method for water absorption of plastics (ASTM D 570-98). Water absorption was measured using a 24 hr immersion test conducted at 19°C. Sample dimensions for both tests were 13 mm x 13 mm x 305 mm.

LCTE was determined using the standard test method for determination of the LCTE of plastic lumber and plastic lumber shapes between -30°F and 140°F (-34.4°C and 60°C) (ASTM D6341-98). Samples were 13 mm x 13 mm x 305 mm and the LCTE was measured in the axial or extruded direction. The actual low and high temperatures used for this test were 26°F and 168°F, respectively.

Flexural properties of the composite samples were determined following the standard test methods for flexural properties of unreinforced and reinforced plastic lumber (ASTM D 6109-97). With the sample depth-to-width ratio less than two, a typical four-point bending test configuration was used. A test jig was constructed in accordance with the above ASTM standard. The load span was 68 mm with a support span of 204 mm. Tested samples had a 13 mm square cross-section with a length of 254 mm. The deflection rate was 6 mm/min. The modulus of rupture (MOR) was calculated as the ultimate strength under flexion, and modulus of elasticity (MOE) was calculated as the slope of the stress-strain curve under flexion, as specified by the ASTM standard.

Shear properties were tested using the standard test method for shear properties of plastic lumber and plastic lumber shapes (ASTM D 6435-99). Samples were tested in double shear in the plane perpendicular to the axis of extrusion. The specified deflection rate of 0.6 mm/min was not possible with available equipment. A deflection rate of 0.75 mm/min was used, as this was the closest realizable strain rate. Samples were tested using a custom-built jig following the specifications in ASTM D 6435-99 standard. The maximum shear strength (MSS) was calculated as the ultimate strength of the samples under shear test.

Parallel compressive properties of extruded samples were determined using the standard test method for compressive properties of plastic lumber and shapes (ASTM D 6108-97). Blanks tested had 13 mm square cross sections and were cut to



25 mm in the axial/extruded direction. Properties were tested along the axis of extrusion on an Instron model 1011 universal testing machine with a loading rate of .03 (mm/mm/min). Platens produced by Instron were used to support the sample. Compressive strength of samples was calculated as the ultimate strength of samples under compression.

The effect of relative proportion of rice hull on physical and mechanical properties of the composite was tested with a general linear model (GLM) procedure in SAS (SAS, 2003). The means of various mechanical properties under different levels of rice hull substitution were compared with Fisher's least significant difference (LSD) analysis. All tests were considered significant at an alpha value of 0.05.

#### Sample Comparison

Mechanical properties of the new composite samples were compared to commercial decking composites, based on the availability of their property data. The two commercial WPCs used were Trex (Trex Company, Inc.) and Rhinodeck (Master Mark Plastics). Properties reported for a 2x6 rice hull-polyethylene composite board produced by Nexwood are also used in comparison (Pacific Lumber Resources). The mechanical properties and standards used for testing these products were obtained from data published on the company's or distributor's websites (Trex Company, Inc. 2005, Master Mark Plastics 2004, Pacific Lumber Resources, 2006). Trex produces a WPC lumber

composed of approximately 40%-50% HDPE and 50%-60% wood fiber (Trex Company, Inc. 2003). Master Mark's Rhino Deck product line contains 30%-50% HDPE, 50%-65% wood flour (Master Mark Plastics, 2004). Nexwood's composition is approximately 60% rice hull and 40% polyethylene (Pacific Lumber Resources, 2006).

#### Results and Discussions:

The lignocellulosic composite having wood flour substituted with rice hull at varying proportion exhibited light weight with specific gravity lower than unity (Table 3). Specific gravity of the new composite varied from 0.95 to 0.99, and as the relative proportion of rice hull in the composite increased, its specific gravity increased. A low specific gravity is preferred for building material applications. When the new composite was compared to commercially available composites, its specific gravity fell within the range of reported values for commercial products. Specific gravities of the new product were all greater than that of both Nexwood and Rhino Deck composites.

Water absorption is an important property for composite materials that are exposed to environmental extremes. A lower water absorption rate is preferred as water absorption can lead to decay and shortened life of cellulosic materials in the composite. Water absorption rate of the new sanded composite samples varied from 1.1 to 2.4 % (Table 3). Mean water absorption for

Table 3: Mechanical properties of the produced lignocellulosic composite material developed from rice hulls, in comparison to two commercial WPC materials.

Composite material	Sp. Gr. at 19/19°C	Water Abs. (%)	LCTE ( $\mu\text{m}/\text{m}^{\circ}\text{C}$ )	MOE (MPa)	MOR (MPa)	MCS (MPa)	MSS (MPa)
Wood	0.960 <sup>ac</sup>	2.2 <sup>a</sup>	54.61 <sup>a</sup>	931 <sup>a</sup>	9.29 <sup>a</sup>	12.2 <sup>a</sup>	5.22 <sup>a</sup>
RH10	0.977 <sup>b</sup>	1.8 <sup>b</sup>	48.85 <sup>a</sup>	1057 <sup>a</sup>	10.76 <sup>a</sup>	11.6 <sup>a</sup>	5.05 <sup>a</sup>
RH20	0.973 <sup>bc</sup>	1.6 <sup>bc</sup>	62.59 <sup>a</sup>	836 <sup>a</sup>	9.86 <sup>a</sup>	12.1 <sup>a</sup>	4.94 <sup>a</sup>
RH30	0.975 <sup>bc</sup>	1.8 <sup>b</sup>	59.79 <sup>a</sup>	1099 <sup>a</sup>	12.39 <sup>a</sup>	15.9 <sup>a</sup>	6.94 <sup>a</sup>
RH40	0.979 <sup>bc</sup>	1.5 <sup>bc</sup>	64.58 <sup>a</sup>	1163 <sup>a</sup>	15.78 <sup>a</sup>	15.5 <sup>a</sup>	6.90 <sup>a</sup>
RH50	0.989 <sup>b</sup>	1.2 <sup>c</sup>	62.28 <sup>a</sup>	793 <sup>a</sup>	12.17 <sup>a</sup>	13.3 <sup>a</sup>	7.53 <sup>a</sup>
Trex	0.91 to 0.95	4.3-sanded 1.7-unsanded	28.98 to 34.56	1207	9.81	12.5	3.87
Rhino Deck	1.08	N/A	35.82	N/A	33.46-edge 21.79-flat	N/A	9.58
Nexwood	1.178	<.8%	48.7	2780	20.2	13.6	N/A
ASTM standard	D6111 <sup>1</sup> D2395 <sup>2,3</sup> D4442 <sup>3</sup> D792 <sup>4</sup>	D570 <sup>1</sup> D1037 <sup>2,4</sup>	D6341 <sup>1,2,3,4</sup>	D6109 <sup>1,4</sup> D4761 <sup>2</sup>	D6109 <sup>1,4</sup> D4761 <sup>2,3</sup>	D6108 <sup>1,4</sup> D198 <sup>2</sup>	D6435 <sup>1</sup> D143 <sup>2,3</sup>

- <sup>1</sup>Rice Hull Samples with Postscript Denoting Hull Content as Percent by mass, <sup>2</sup>Trex, <sup>3</sup>Rhino Deck, <sup>4</sup>Nexwood.
- <sup>a,b,c</sup> Different letters in a column indicate significant difference (P < 0.05) between the treatments.

each treatment dropped from 1.8 to 1.2% as the rice hull content was increased from 10 to 50%. This was significantly lower than the water absorption rate of samples with no rice hull, which was 2.2%. All the samples tested in this experiment had superior water absorption property compared to the Trex product. The reported water absorption rate for Trex is 1.7% for unsanded and 4.3% for sanded samples. The reported value for Nexwood was less than .8%. Our samples did not contain any water phobic additives that are commonly used in commercial samples to decrease water absorption. The incorporation of water phobic additives in WPC can further reduce water absorption substantially. Additionally, water absorption for sanded products is significantly higher than unsanded samples as the plastic coating protects the sample from water in unsanded samples.

Linear coefficient of thermal expansion expresses the temperature stability of the composite material. This is an important property if the material is intended for use in outdoor environments, as the expansion and compression under extreme weather events would lead to failure of joints. The mean LCTE value of the samples under different rice hull composition varied from 49-65%, which was much higher than reported values of Trex and Rhino Deck and was somewhat higher than that reported for Nexwood (Table 3). A comparison with the samples extruded under similar conditions with no rice hull indicated a high LCTE for these samples as well. A statistical analysis of means with Fisher's least significant difference (LSD) indicated no difference in LCTE between different rice hull treatments. For these reasons it is presumed that the differing production processes may have affected the LCTE.

The MOE, determined using the 4-point bending tests indicates the toughness of the composite material. MOE for the new composite varied from 793 to 1163 MPa. The MOE of the new composite is lower than both Trex (1207 MPa) and Nexwood (2780 MPa). However, it is believed that this property can be increased significantly by using higher extrusion pressure than were achievable with the arbor press used. This is also supported by the fact that the composite samples produced with the same

process with no rice hull also showed a low MOE of 931 MPa. Analysis of mean MOE for each treatment showed no significant difference in MOE between various rice hull treatments.

The ultimate strength properties (ultimate strengths in parallel compression, transverse shear, and pure bending) of the composite samples were comparable to reported values for commercial WPC products (Table 3). The strength properties of the new composite showed numerical increase with increasing proportions of rice hull. The average MOR of the new composite samples varied from 9.86 to 15.78 MPa for samples with 9.6-48% rice hull, which was significantly higher than the Trex MOR of 9.8 MPa. Mean parallel compressive strengths varied from 11.6 to 15.9 MPa. The compressive strength of both Trex and Nexwood were within the range for the new composite. Shear strength showed significant increases with the amount of RH average values varying from 5.05 to 7.53 MPa. The shear strength of the new composite was superior in its shear strength properties compared to Trex (3.5 MPa). It, however, must be noted that shear strength presented by both manufacturers are in the longitudinal plane whereas shear strength tests of RH samples were conducted in the transverse plane.

Analysis of variance (ANOVA) on measured properties of the new composite samples with respect to the relative proportion of rice hull showed that percentage rice hull significantly influenced both specific gravity and water absorption (Table 4). Approximately 55% of variability in specific gravity and 78% of variability in water absorption was explained by the relative proportion of rice hull in the new composite.

It is clear from the comparison between the physical and mechanical properties of the new composite with various proportions of rice hull and the commercial products that lignocellulosic composites made from rice hull can be reasonably be offered as an alternative for the currently available WPCs in the market. Most of the market available WPCs are used for non-structural applications where high strength properties are not required. The new composite we have developed can be used for

*Table 4. Effect of rice hull substitution on various physical and mechanical properties of a lignocellulosic composite with 48% plastic. The rows with bold-italic numbers indicate that substitution with RH significantly influenced the specific property.*

Property	Model DF	Mean	RMSE	p-value of treatment	R <sup>2</sup> value
Specific gravity	5	<b>0.98</b>	<b>0.0099</b>	<b>0.0584</b>	<b>0.55</b>
Water absorption	5	<b>1.68</b>	<b>0.22</b>	<b>0.0013</b>	<b>0.78</b>
LCTE	5	31.21	8.19	0.4773	0.28
MOE	5	140574.1	54533.36	0.8126	0.15
MOR	5	1663.33	738.51	0.7816	0.17
Shear Strength	5	874.97	269.75	0.3545	0.33
Compressive Strength	5	1888.64	595.62	0.7607	0.18

non-structural applications such as decking, fencing, flooring and OEM applications. It has to be noted that we have employed an extruder developed inhouse to manufacture the samples. The pressure and temperature controls of this extruder were not as accurate as that of a commercially available extruder. Similarly, the extrusion pressure was much lower than that of a commercial extruder. Therefore, it is possible that samples generated with similar proportion of rice hull in a commercial extruder will have significantly high strength properties due to the high proportion of silica in the hull. Similarly, it may also exhibit higher specific gravity. The major drawback of using rice hulls in lignocellulosic composite materials is its high abrasiveness, which increases the wear of the extruder.

### Conclusions:

A study was conducted to test whether raw or parboiled rice hulls could be used as a lignocellulosic substitute in wood-plastic composites. A typical WPC composition of 48% wood, 48% plastic and 4% Zinc stearate was used for the study. Wood flour was substituted by RH in increments of 20% by weight, varying from 0% (no RH) to 100% RH. The composites exhibited low specific gravity and water absorption, two critical properties for WPCs. The rice hull in the composite tended to increase the specific gravity and decrease water absorption. The new composite was superior to commercial products in its water absorption, mainly due to the hydrophobic nature of rice hull. Properties such as specific gravity, LCTE, MOR, compressive strength and shear strength of the new composite was comparable to that of commercial WPC products. With better temperature and pressure control in the extrusion process, the properties of the new composite could improve significantly. In conclusion, rice hull is a renewable and viable alternative source of lignocellulosic material, and could be successfully used in lignocellulosic-plastic composites. While handling and processing of the rice hull will be similar to wood flour, its use could prove desirable to WPC manufacturers in rice producing regions.

### Literature Cited:

- The Agrilectric Companies. 2005. Plant Schematic. The Agrilectric Companies, Lake Charles, La. Available at <http://www.agrilectric.com/info.asp?caid=2&cid=17>.
- Bronzeoak, Ltd. 2003. Rice Husk Ash Market Survey. Bronzeoak, Ltd., Caterham, UK. Available at <http://www.dti.gov.uk/renewables/publications/pdfs/exp129.pdf>
- Cramer, G. L., E. J. Wailes, and K. B. Young. 1991. Arkansas. Special Report 152. Arkansas Agricultural Experiment Station- Division of Agriculture, Fayetteville, AR.
- Freedonia Group Inc. 2004. Composite and Plastic Lumber. Freedonia Group Inc., Cleveland, Ohio.
- Hunter, A. M. 1991. Utilization of annual plants and agricultural residues for the production of pulp and paper products. Nonwood Plant Fiber Pulping Progress Report 19, TAPPI Press, Atlanta, Georgia.
- Juliano, B. O. 1985. Rice hull and rice straw. In: Rice: Chemistry and Technology, pp.689-755. edited by B. O. Juliano. Westport, CT: AVI.
- Luh, B. S. 1991. Rice Vol. II Utilization. Van Nostrand Reinhold, New

York, NY. pp. 269-294.

Master Mark Plastics. 2004. Rhino deck physical and mechanical properties. Master Mark Plastics, Albany, MN. Available at [http://www.mastermark.com/rhino\\_deck/properties.asp](http://www.mastermark.com/rhino_deck/properties.asp).

Master Mark Plastics. 2004. Material Safety Data Sheet No: BRN-01. Master Mark Plastics, Albany, MN. Available at [http://www.mastermark.com/rhino\\_deck/pdf/msds.pdf](http://www.mastermark.com/rhino_deck/pdf/msds.pdf).

Miller, D. E., and R. A. Eisenhauer. 1982. Agricultural byproducts and residues. In: CRC Handbook of Processing and Utilization in Agriculture. Vol. 2 Part I Plant Products, pp. 691-708. edited by I. A. Wolf. Boca Raton, FL: CRC Press.

Morton, J., J. Quarmley, and L. Ross. 2003. Current and emerging applications of natural and woodfibre-plastic composites. In: Proc. Seventh International Conference on woodfiber-plastic composites, pp.3-6. Forest Product Society, Madison, WI.

National Academy of Sciences. 1971. U.S. and Canadian feeds. In: Atlas of Nutrition Data. National Academy of Sciences, Washington, D. C.

Saha, B. C., L. B. Iten, M. A. Cotta, and Y. Wu. 2004. Rice hull as substrate for production of fuel ethanol. In: Proc. The 33rd Annual Meeting Of The United States-Japan Cooperative Program In Natural Resources (Unjr), p. 181-185. edited by J.P.Cherry and A.E. Pavlath. December 11-18, 2004, Honolulu, Hawaii..

SAS. 1999. SAS Users Manual, SAS Institute Inc. Cary, NC.

Trex Company, Inc. 2005. Trade professionals - physical and mechanical properties. Trex Company Inc., Winchester, VA. Available at [http://www.trex.com/Universal/technical\\_info/properties.asp](http://www.trex.com/Universal/technical_info/properties.asp).

Trex Company, Inc. 2003. Material safety data sheet MSDSNAT-01. Trex Company Inc., Winchester, VA. Available at [http://www.trex.com/Universal/technical\\_info/MSDS-Natural.pdf](http://www.trex.com/Universal/technical_info/MSDS-Natural.pdf).

Oliver, P.A. 2004. The rice hull house. The International Journal of Straw Bale and Natural Building. 47. Available at <http://www.thelaststraw.org/backissues/articles/Rice%20Hull%20House.pdf>

Pacific Lumber Resources. 2006. About nexwood. Pacific Lumber Resources., Lake Oswego, OR. Available at <http://www.nexwoodnw.com/about-nexwood/>

Pacific Lumber Resources. 2006. Nexwood Test Results. Pacific Lumber Resources. Lake Oswego, OR. Available at <http://www.nexwoodnw.com/pdf/handbook/15-test-results.pdf>

Primenergy, L.L.C. 2006. Case Study: Rice hulls to process heat and steam. Primenergy, L.L.C. Tulsa, OK. Available at [http://www.primenergy.com/Projects\\_detail\\_jonesboro.htm](http://www.primenergy.com/Projects_detail_jonesboro.htm)

Virginia Division of Energy. 2006. Virginia Energy Saver's Handbook. Available at <http://www.mme.state.va.us/de/hbchap5.html>

Winandy, J.E., N.M. Stark, and C.M. Clemons. 2004. Considerations in recycling of wood-plastic composites. In: Proc. 5th Global Wood and Natural Fibre Composites Symposium, p. 9. April 27-28, 2004, in Kassel, Germany.

### Mentor comments:

Dr. Skreekala Bajwa recommended her student's work highly. In her letter in support of publication of Mr. Bourne's work, she said,

I was excited to nominate the article 'Evaluation of rice hulls as a lignocellulosic substitute in wood plastic composites' for publication in Inquiry. This article provides a solution to an environmental problem faced by the crop processing industry in Arkansas. Although it is known that rice hulls could substitute for wood in wood plastic composites, there is no

published data on the properties of such composites. This research provides valuable information to researchers and the wood plastics composite industry while providing a value-added product to the rice processing industry for their waste stream. The wood plastic composite industry is a fast growing industry with significant presence in Arkansas.

I would attribute the success of this project primarily to Jack's motivation, and to a well-planned research program. Jack was a quick learner and a pleasure to work with. He took courses on mechanics of materials, and mechanical design at the same time he was working on the project. He was excited to see practical applications of the theory he was learning in those courses. I was impressed by his quick intellect, sincerity towards the project, and the extra measures he took to ensure reliability of data generated. He performed Master's level research, and it is the first time in my five-year career that I have come across a student of such high caliber.

## ULTRAFILTRATION MEMBRANE PROCESS FOR PYROGEN REMOVAL IN THE PREPARATION OF WATER FOR INJECTION (WFI)

By Thet N. Soe

Ralph E. Martin Department of Chemical Engineering

Faculty Mentor: Professor Robert Cross

Ralph E. Martin Department of Chemical Engineering

### Acknowledgements:

First of all, I would like to thank to Professor Robert Cross from Ralph E. Martin Department of Chemical Engineering for giving me the opportunity to work on this project. I gratefully acknowledge and thank him for his advice, supervision, and generous help throughout the project. Additionally, it is a great pleasure to acknowledge the assistance I have received from a graduate student, Liming Song, in using new equipment in the lab. Last but not least, I would like to express my sincere gratitude to Honors College for funding this project. Without the help of Brenda Robertson, the materials required for the experiment would not have arrived in time. Again, I would like to thank all those people that made this project an enjoyable experience for me.

### Abstract:

An artificial kidney process based on ultrafiltration (called hemodiafiltration) more effectively removes blood toxins, particularly those of higher molecular weight, than conventional dialysis. In hemodiafiltration, replacement liquid must be added directly to the patient's blood to replace liquid filtered out of the blood in the ultrafiltration cartridge. This replacement liquid is made up of "water for injection" (WFI) and other components such as salts and glucose. WFI is produced either by distillation or reverse osmosis and has a cost of \$1.00 per liter. A patient with kidney problems would require three hemodiafiltration treatments per week, each requiring 70-80 liters of WFI. Since the target cost of a treatment is less than \$100, the cost of WFI alone, \$70-\$80, makes the process not economically feasible. In order for the hemodiafiltration process to be widely used and to be affordable in third world countries, a system that produces WFI at a low cost is needed. A process to replace distillation or reverse osmosis must be capable of removing pyrogens. Some information is available in the literature indicating that ultrafiltration membranes can remove endotoxins, the main constituent of pyrogens; however, no studies have been done to establish the type of ultrafiltration membrane that gives optimal removal of endotoxins.

Ultrafiltration is known to be a much less expensive process than distillation or reverse osmosis. It is thoroughly estimated that WFI produced by Ultrafiltration System would cost only 25 cents per liter. The purpose of this experimental work is to determine the type of ultrafiltration membrane that effectively removes endotoxins with an efficient flow rate. Regenerated cellulose and polyethersulfone membranes with various molecular-weight-cut-offs were evaluated to determine the endotoxin rejection and flux rates of the membranes. A stirred cell experiment was performed as a short-term test, using disc membranes with a diameter of 76 mm and three types of feed solutions. The *Limulus Amebocyte Lysate* (LAL) Gel Clot test was performed to measure the concentration of pyrogen in the filtrates. The best candidate from the stirred cell experiment was tested in a Hollow Fiber Cartridge Ultrafiltration System over a longer time period.

The results showed that a polyethersulfone membrane with a molecular-weight-cut-off (MWCO) of 10,000 rejected endotoxin to below the US Pharmacopeia (USP) limit of endotoxin content, 0.25 EU/ml, with the best flux rate. Since polyethersulfone membranes were not available in a cartridge form, the membrane with the closest molecular structure, a polysulfone membrane, was tested in a Hollow Fiber Cartridge Ultrafiltration System. The results showed that the polysulfone membrane cartridge rejected endotoxin content from 625 EU/ml to less than 0.25 EU/ml consistently over a week-long test period. In addition, the flux rate remained constant at 129 L/m<sup>2</sup>/hr. Thus, the polysulfone membrane of 10,000 MWCO can be used in an ultrafiltration system to produce Water for Injection (WFI).

### Introduction:

Water for injection (WFI) is used to make up the solution used to replace the fluid lost from blood during hemodiafiltration, an artificial kidney process. In the regulations of US Pharmacopeia (USP), the primary purification process in manufacturing WFI is required to be either distillation or reverse osmosis. Either of those processes produces WFI at a cost of \$1 per liter. In a hemodiafiltration session, 70 to 80 liters of WFI is used, and the

most common regimen is three times a week. A patient uses nearly 15,000 liters of WFI in a year. Due to the high cost of WFI, the hemodiafiltration process is not affordable to all patients. In this work, ultrafiltration membranes were studied for removal of endotoxin, the main component of pyrogen, in order to substitute for current costly purification methods.

### Background:

Dialysis is a routine process to treat patients with kidney failures. A regular dialysis method cannot, however, remove middle and high molecular weight blood toxins. The hemodiafiltration method, which is a modification of the regular dialysis method, can be used to remove middle and high molecular weight blood toxins.

In the hemodiafiltration process, a patient's blood flows through a hemodiafiltration artificial kidney, which removes middle and high molecular weight blood toxins by increasing filtration. A saline solution made up with water for injection, also known as "reconstituting fluid," is infused either into the inlet or the outlet of the hemodiafiltration kidney to replace the fluid lost from the blood.

On average, a patient with kidney failure uses 70 to 80 liters of WFI in a hemodiafiltration session, totaling nearly 15,000 liters in a year. In order for the hemodiafiltration process to be widely used and to be affordable in third world countries, a system that produces WFI from tap water at a low cost is necessary.

### Literature Review:

In a normal dialysis procedure, a significant amount of endotoxins (ET) as well as other pyrogenic substances are detected. Endotoxins are fragments of dead bacteria. Dialysate containing endotoxins higher than the limit can cause endothelium damage, arteriosclerosis, and inflammatory problems such as amyloidosis. The major components of endotoxins are lipopolysaccharides (LPS) from gram-negative bacteria walls. The LPS can be divided into three substructures: a hydrophobic and high conservative structure formed by lipid A (non-polar), a core oligosaccharide conservative part, and a variable heteropolysaccharide surface structure. The last two parts can vary largely in structures and sizes, depending on bacterial species. This leads to a wide range of molecular weights, 3-25 kDa (average 10-12 kDa), with a variety of biological activities. Pyrogens are very heat stable compounds and cannot be eliminated by autoclaving or microfiltration.

The Association of Advancement of Medical Instrumentation (AAMI) estimates that approximately 75 million endotoxin units (EU) of ET and 30 billion bacteria passed through the "water side" of the dialyzer during one year of hemodialysis treatments. According to its survey, 19-35% of water samples from dialysis centers had bacterial counts above

the standards, and 6 % of dialysate samples exceeded the AAMI endotoxin limit (5 EU/ml). It was highly expected that the regular use of sterile and endotoxin-free dialysate would help decrease the cardiovascular diseases and mortality rate of patients undergoing hemodialysis and, of course, would be absolutely essential for patients undergoing hemodiafiltration.

Few methods to produce WFI without using distillation or reverse osmosis can be found in the literature. Reti and Benn (U.S. Pat. No. 4,610,790) showed a method of using a plurality of filtration and deionization steps producing sterile water. Harris (U.S. Pat. No. 3,959,128) used non-ionogenic hydrophobic synthetic polymers to absorb endotoxin from biological fluids.

In June 1993, *Kidney Intl Journal* described an experimental circuit that was used to ultrafilter a bicarbonate dialysate contaminated (5 to 48 EU/ml) by a *Pseudomonas Aeruginosa* for 240 hours with the flow rate of 500ml/min. In some research, ultrafiltration membranes were observed to remove endotoxin; however, no studies have been done to establish the type of ultrafiltration membrane that gives optimal removal of endotoxin.

In this experiment, various types of ultrafiltration membranes were studied to remove endotoxins. The stirred cell test was performed as a short-term test with cellulose acetate membranes and polyethersulfone membranes with a diameter of 76 mm, each having specific molecular weight cut-offs (MWCO). The LAL gel clot method was used to determine the endotoxin level in the filtrates. Three types of feed solutions with different endotoxin concentrations were used to study the endotoxin concentration with respect to the rejecting capability of the membranes. Flux rate versus concentration of the feed solutions was measured. The membrane that gave the best result from the short-term test was further tested in a Hollow Fiber Cartridge Ultrafiltration system as a long-term test. A feed solution with a high endotoxin challenge level was used in the long term process to assure the endotoxin rejecting capability of the membrane.

### Materials and Methods:

Extreme care was necessary in all tests to avoid contamination problems. Safety gloves and goggles were worn during the entire experiment. Water used in all of the experiments was pyrogen-free water known as Water for Irrigation from Abbott Laboratories.

#### *Stirred Ultrafiltration Cell (UF) test*

A Millipore stirred UF cell, model no. 8400, with a volume of 350 ml was used in the experiment. Disc membranes with a diameter of 76 mm were also from Millipore Corporation.

First of all, the stirred cell was disassembled, and the parts were submerged in a beaker containing 95% ethanol solution for 15 minutes to remove all pyrogens, residues, and bacteria. Then, the stirred cell was thoroughly rinsed with water to remove all the chemicals and contaminants stuck on the cell. As the membranes

were not sterilizing-grade, a sanitizing method was necessary to remove all the particles and residues on the surface and in the pores of the membranes. Each membrane was dipped in a plate containing 200 ppm of sodium hypochlorite solution before the experiment. Then, the membrane was also rinsed with water to remove the chemicals. After sanitization, in order to avoid scratching and contaminating the surface, only the edge of the membrane was held and the membrane was placed in the membrane holder of the cell. An O-ring was put on top of the membrane, and the membrane holder was screwed firmly to the base of the cell. The parts of the cell were put together, and the cell was filled with 350 ml of feed solution. The filled cell was placed in its retaining stand, and the cell, together with its retaining stand, was placed on a magnetic stirrer. The stirrer was turned on enough for the stirrer in the cell body to keep moving to reduce membrane fouling. Then, with the pressure-relief knob in the vertical (closed) position, the inlet line was connected to the pressure-regulated nitrogen gas cylinder with a tube. It is essential to be aware that operating the cell without its retaining stand can result in cap popping off and splattering contents during pressurization. With the pressure valve closed, the pressure gauge was set to 60 psi. Increasing pressure above a critical point may result in a lower flux rate due to compact layer of retained materials on the membranes. It was also necessary to be aware that the pressure limitation of the stirred cell was 75 psi. While holding the cell steady on the magnetic stirrer, the pressure valve was turned on, and timing was started with a stop watch to study the flux rate. The first 200 ml of the permeate was drained out from the filtrate exit tubing into a beaker to not collect residues stuck in the cell and tubing. Using a sterile, non-pyrogenic 45-ml centrifuge tube, about 5 ml of the sample was collected. Timing was stopped when the cell was empty. Then, another 350ml of feed solution was poured into the cell and followed the same steps for flux rate study. Totally, four runs using a total of 1,400ml of feed solution were conducted on each membrane in order to get enough data for the flux rate study. Flux rate was determined by the following formula.

Net diameter = diameter of membrane – diameter of O-ring

Flux rate = volume of water (L) / (Net area of the membrane (m<sup>2</sup>) \* time (hr))

The endotoxin concentration in the permeate was measured by using the LAL gel clot test.

#### *Pyrosate Limulus Amebocyte Lysate (LAL) Gel Clot test*

All the materials required for the LAL test were purchased from Associates of Cape Cod Inc. A dry-bath incubator was turned on and calibrated to 37 °C following the instruction in the manual. It was necessary to be aware of a few things. When the temperature first reached 37°C, the temperature was not stable yet. It would be noticed that the temperature would go beyond 37

°C, go back down below 37 °C, and then go back to 37° C. When the temperature stayed at 37± 1 °C, the incubator was ready to use.

LAL gel clot kits used in this study had the sensitivity of 0.25 EU/ml. Dilution tubes made of flint (soda lime) glass, LAL reagent water (LRW), a four-channel digital stop watch, and non-pyrogenic disposable pipettes were also used for the gel clot test.

Limulus amebocyte lysate is an aqueous extract of blood cells (amebocytes) from the horseshoe crab, *Limulus polyphemus*. Gram negative bacteria cause Limulus blood to clot. It was later determined that clotting was initiated by a unique component of the bacterial cell wall called endotoxin or lipopolysaccharide.

The endotoxin concentration in the solution before and after the experiment was determined. Good labeling skills and sanitization methods are highly recommended for this test.

First of all, four pairs of SPL (blue capped sample tube) and PPC (red capped positive control tube) were put in two rows in a rack. The stoppers were removed and the tubes were put back in the rack. Four dilution tubes were placed in another rack to perform a series of dilutions. Using a non-pyrogenic individually wrapped pipette, each of the two dilution tubes were filled with 1-ml of sample solution from the centrifuge tube. The pipette was labeled '1' and was temporarily put back into the package for further use. Using another new pipette, the second dilution tube containing 1-ml of sample solution was added with 1-ml of LRW water to make the concentration half. The pipette was labeled 'L' and put it back into the package. Then, the second dilution had a total of 2 ml solution. Using a new pipette, a solution of 1 ml was transferred from the second tube into the third tube and labeled the pipette '2'. Using the pipette labeled 'L', LRW water of 1-ml was added to make the concentration one-fourth. Following the same procedure, the concentration in the fourth tube was made one-eighth concentration.

After the dilution task was performed, the pipette labeled '1' was used to transfer 0.5 ml of solution from the first tube into a SPL tube. The tube was shaken vigorously for 15 to 20 seconds to ensure thorough mixing. The contents should dissolve in the solution in this period. Failure to dissolve adequately can produce unsatisfactory results. After the contents had dissolved, 0.25 ml of solution from SPL tube was transferred into a PPC tube for positive control test using the same pipette labeled '1'. The same procedure was followed to transfer the solution from the dilution tubes into SPL and PPC tubes respectively.

The SPL and PPC tubes were immediately transferred into a dry-bath incubator and incubated at 37±1 °C for the period mentioned in the LAL manual. The incubating period can be different for different packages, but generally it is 25 minutes. When the time was up, each tube was removed from the incubator, and the results were interpreted. A positive test was indicated by



the formation of a gel that does not collapse in a 180 degree inversion. A negative test was indicated if the content was still in aqueous state or the gel clot collapsed when the tube was inverted 180 degrees gently.

The PPC tubes must be test positive to rule out false negative results. If the positive control was negative, the SPL tube was positive, the specimen was believed to be interfering with the LAL test, and the test should be redone.

The endpoint of the test was defined as the least concentration of endotoxin to give a positive test. The following method of determining the concentration of endotoxin was used. For example, the endpoint was assumed to be at one-fourth dilution,

Specimen Dilution	Test Result
Undiluted	+
1/2	+
1/4	+
1/8	-

Pyrotell sensitivity (\*), in this case 0.25 EU/ml, was multiplied by the reciprocal of the dilution at the endpoint to determine the concentration of endotoxin.

$$\text{Endotoxin Concentration} = (*) (4/1) = 0.25 \text{ EU/ml} * 4 = 1 \text{ EU/ml}$$

If the negative result was not indicated in all four SPL tubes, more dilution tubes were prepared until one sample solution gave the negative result in a SPL tube.

Three types of feed solutions were used in this experiment: tap water, Milli Q water, and pond water. Milli Q water was obtained from Milli Q water filtration system. To challenge the resistance of the membrane, pond water, which contained high endotoxin concentration, was used. The endotoxin concentration of tap water and Milli Q water were determined by LAL Gel clot method in the lab. Due to limited material, pond water was sent to Associates of Cape Cod to determine the endotoxin concentration.

#### *Hollow Fiber Cartridge Ultrafiltration (UF) System*

The Hollow fiber Ultrafiltration System was from Koch Membrane System Inc. The cartridge studied in this experiment was also from Koch Membrane System Inc. The system was made of three basic components: a 12 L process tank, a pump, and a cartridge. As no cooling system was used in the system, the temperature of the feed solution would increase slightly as the fluid was pumped. However, it would not affect the results of the experiment since pyrogens are heat stable compounds.

First of all, the system was sanitized with the following three steps of cleaning cycle. After each step, the solution was drained out of the system by opening V3 as shown in appendix

A.

1. Milli Q water, 5L, was flushed through the entire system for 20 minutes.
2. Sodium Hypochlorite, 5L, with a concentration of 200ppm was circulated in the system for 20 minutes.
3. Water for Irrigation, 5 L, was run in the system for 20 minutes.

Each step of the cleaning cycle had its own significance. The Milli Q water rinsing cycle was to remove residue in the system. As hypochlorite is an oxidizing agent for membranes, sodium hypochlorite cleaning cycle was used to remove bacteria and pyrogen lodged in the pores of the membrane. Water for Irrigation cycle was to remove the chemicals and materials remaining in the cartridge.

Since the chlorine resistance of the polysulfone membrane was 200 ppm for short term sanitation, it was suggested to consult the vendor if longer sanitation period was necessary.

To make sure the system was completely cleaned, 4 L of water for irrigation was circulated in the system. Before the sample was collected, the tip of outlet tubing to which clamp no. 1 was attached was disinfected with 95 percent ethanol solution by dipping in a beaker. To remove the chemical stuck on the tubing, the tubing was dipped and rinsed with water for irrigation. When the sample was collected, clamp 2 (C2) was closed and C1 was open. The first 300 ml of permeate was drained out, and then the sample was collected in a 45-ml non-pyrogenic centrifuge tube. The endotoxin concentration in the permeate was determined by the gel clot test. Only when the permeate measured less than 0.25 EU/ml was the system believed to be clean and ready to use.

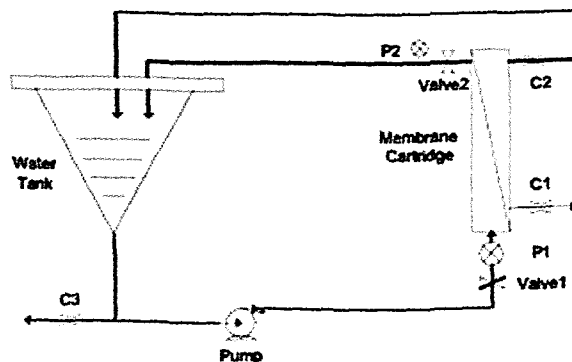


Figure 1: Schematic Diagram of the Hollow Fiber Ultrafiltration System.

A schematic diagram of the hollow fiber cartridge ultrafiltration system is shown in Figure 1. Pond water 5 L was poured into the tank, and the pump was started with downstream valve (V2), Clamp2 (C2) open, and C1, C3 closed. Then, the upstream valve (V1) was turned on gradually until the upstream pressure (P1) reached 10 psi. The downstream valve (V2) was closed slowly until the pressure (P2) was close to 15 psi. The upstream pressure (P1) increased as the downstream pressure (P2) was raised. By adjusting the two pressure valves, P1 was

read 25 psi and P2 15 psi. The transmembrane pressure of 10 psi (the difference between P1 and P2) was held constant for the entire experiment. It was highly recommended not to use the pressure more than 25 psig since the pressure limitation of the hollow fiber cartridge is 25 psig. The system was operated for 8 days continuously. The sample was collected every 24 hour, and the endotoxin concentration was determined by the gel clot test. The feed solution was drained out by opening C3 after collecting sample at the end of the operation.

The operation was immediately followed by the cleansing maintenance so as not to let the membrane dry out in contact with feed materials. Sodium hypochlorite, 5 L, with the concentration of 200ppm was pumped through the system for 20 minutes to remove the feed materials in the pores of the membrane, and then water for irrigation was circulated in the system for 20 minutes to wash away all the chemicals and residues.

### Results and Discussion:

The results from the Stirred Ultrafiltration Cell Test and the Hollow Fiber Ultrafiltration System are presented below.

#### Stirred Ultrafiltration Cell (UF) test

LAL gel clot test showed that Milli Q water had endotoxin content of 1 EU/ml and tap water of 16 EU/ml. As previously mentioned, according to the report from Associates of Cape Cod, pond water had endotoxin content of 625 EU/ml.

Totally, four runs were conducted on each membrane with the same type of feed solution, and permeate sample was collected after each run. The LAL test was performed to determine the endotoxin concentration of the permeate, and the results are shown in table 1.

#### Regenerated Cellulose membranes

As shown in Table 1, none of the regenerated cellulose membranes rejected the endotoxin concentration to less than 0.25 EU/ml. It was observed that endotoxin rejection declined

MWCO	Milli Q (= 1 EU/ml) Endotoxin con. (EU/ml)	Tap water (= 16EU/ml) Endotoxin con. (EU/ml)	Pond water (= 625 EU/ml) Endotoxin con. (EU/ml)
100,000	0.25	0.25	$\geq 0.25$
30,000	0.25	0.25	0.25
10,000	0.25	$\geq 0.25$	$\geq 0.25$

Table 1. Endotoxin concentrate in the permeate from regenerated cellulose membranes.

slightly with the pond water, which contained high challenge level of endotoxin. It suggested that the endotoxin rejecting capability of the regenerated cellulose membranes decreased with the high level of endotoxin.

As seen in Figure 2 and Figure 3, the flux rates of regenerated cellulose membranes decreased with respect to the endotoxin

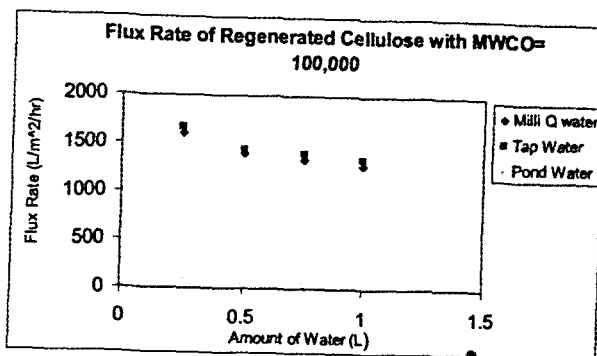


Figure 2. Flux rates of regenerated cellulose membrane with MWCO = 100,000

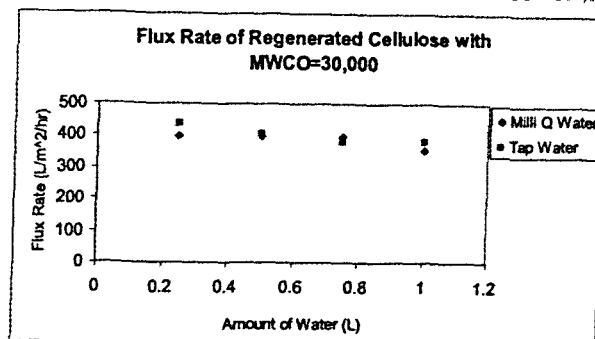


Figure 3. Flux rates of regenerated cellulose membrane with MWCO = 30,000

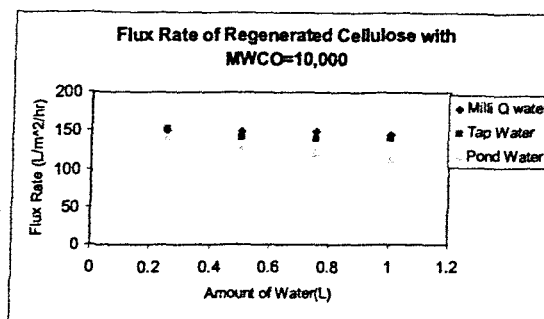


Figure 4: flux rates of regenerated cellulose membrane with MWCO = 10,000

concentration of the feed solution as well as the volume of the solution. In Figure 2, the flux rate of tap water declined slightly compared to that of Milli Q water as the endotoxin concentration was higher in tap water. For pond water, the flux rate was much lower than that of the other feed solutions. It was presumed that due to the high concentration of solids in the feed solution, fouling of membranes increased and the flux rates declined

MWCO	Milli Q (= 1 EU/ml) Endotoxin con. (EU/ml)	Tap water (= 16EU/ml) Endotoxin con. (EU/ml)	Pond water (= 625 EU/ml) Endotoxin con. (EU/ml)
500,000	0.25	> 2	> 16
300,000	0.25	> 2	1
100,000	N/A	< 0.25	0.25
50,000	N/A	< 0.25	0.25
30,000	N/A	< 0.25	0.25
10,000	N/A	< 0.25	< 0.25

Table 2: Endotoxin concentration in the permeate from polyethersulfone membranes

For regenerated cellulose membrane with MWCO = 10,000, the same phenomenon was observed. The flux rate of Milli Q water was the highest, followed by that of tap water, and by pond water. From this observation, it is inferred that membrane fouling was significant with disc membranes. The flux rates decreased a lot from 0.25L of solution to 0.5 L, and formed a parabolic curve after 0.5 L of the feed solution.

#### Polyethersulfone membranes

From Table 2, it was observed that endotoxin rejection of polyethersulfone membranes with MWCO 500,000 and MWCO 300,000 could not meet US Pharmacopeia (USP) limit (<0.25 EU/ml). For MWCO 100,000, MWCO 50,000, MWCO 10,000, endotoxin of the feed solution was rejected to less than 0.25 EU/ml. However, with high challenge level of feed solution (pond water), the gel clot test showed 0.25 EU/ml. The polyethersulfone with MWCO 10,000 showed the most satisfactory results rejecting endotoxin below 0.25 EU/ml in permeate of tap water and pond water.

Flux rates of polyethersulfone membranes were plotted in Figures 5, 6, and 7. Due to limited numbers of membranes, only one type of solution was tested for flux rate study. The flux rates declined with respect to the volume of water and formed a parabolic curve after 0.5 L of feed solution as observed previously.

Upon comparison between regenerated cellulose and polyethersulfone membranes, endotoxin rejection of polyethersulfone was more reliable than that of regenerated cellulose membranes. Permeate from regenerated cellulose membrane with MWCO = 10,000 resulted in 0.25 EU/ml, but polyethersulfone membrane with MWCO = 10,000 rejected endotoxin to less than 0.25 EU/ml. The flux rate of polyethersulfone membrane was also nearly four times higher than that of regenerated cellulose membranes. Due to limited mechanism in manufacturing of membranes, the flux rate of one membrane could not be exactly the same as another membrane of the same type. However, flux rate of polyethersulfone membranes were apparently higher than regenerated cellulose membranes.

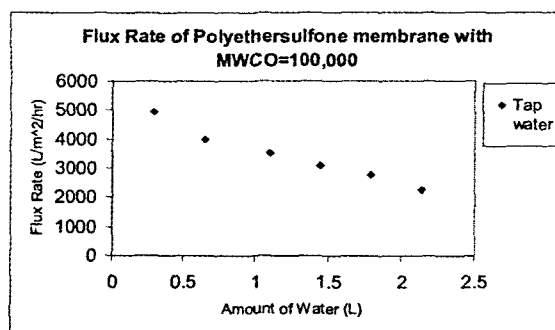


Figure 5: Flux rates of polyethersulfone membrane with MWCO = 100,000.

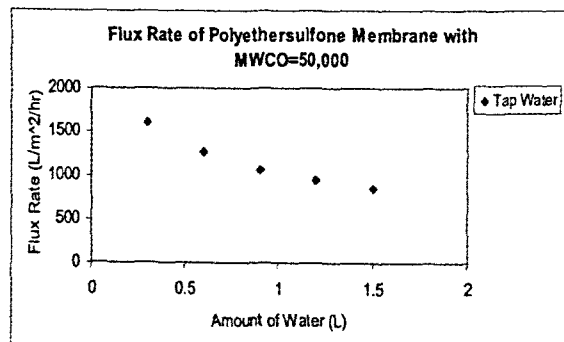


Figure 6: Flux rates of polyethersulfone membrane with MWCO = 50,000.

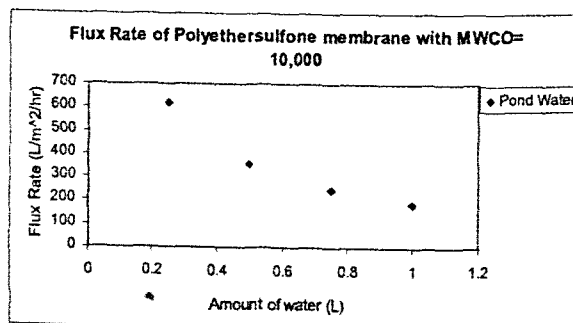


Figure 7: Flux rates of polyethersulfone membrane with MWCO = 10,000.

Therefore, a polyethersulfone membrane with MWCO = 10,000 was chosen to test in Hollow Fiber UF system for a long-term period. Due to limited availability of the membrane in the market, polyethersulfone membrane was not available in cartridge form. The membrane with the closest molecular structure, polysulfone membrane cartridge, was chosen for the Hollow Fiber UF system experiment.

#### *Hollow Fiber UF system experiment*

The results showed that polysulfone cartridge with MWCO 10,000 rejected the endotoxin concentration from 625 EU/ml to less than 0.25 EU/ml. The rejection capability of the cartridge did not decline during the eight-day period. The flux rate stayed the same at 129 L/m<sup>2</sup>/hr. In the hollow fiber cartridge system, fouling of membrane was not observed, and the same flux rate was observed for all the entire period. It can be concluded that polysulfone membrane can reject endotoxin meeting USP standard, and its rejecting capability did not decline in the continuous operation for the eight day period.

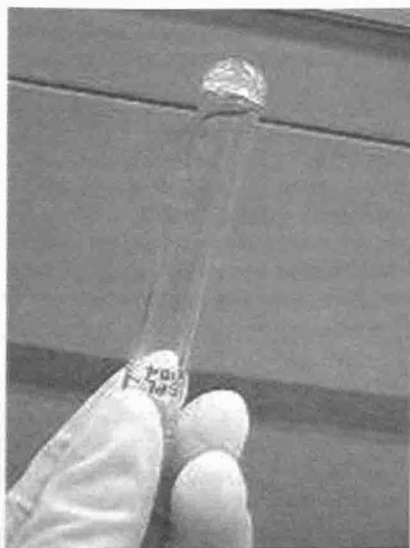
#### **Conclusion:**

Results from short-term experiments showed that regenerated cellulose membranes (MWCO 100,000; MWCO 30,000; MWCO 10,000) cannot reject endotoxin to meet the US Pharmacopeia standard (<0.25 EU/ml). The endotoxin rejection of Polyethersulfone membranes (MWCO 500,000 and MWCO 300,000) is not below 0.25 EU/ml. Polyethersulfone membranes (MWCO 50,000 and MWCO 30,000) can reject endotoxin in tap water to less than 0.25 EU/ml. However, with high level of endotoxin, the membranes cannot meet the standard. For polyethersulfone membrane with MWCO 10,000, the endotoxin measured less than 0.25 EU/ml in the permeates. The flux rates of polyethersulfone membranes were about four times higher

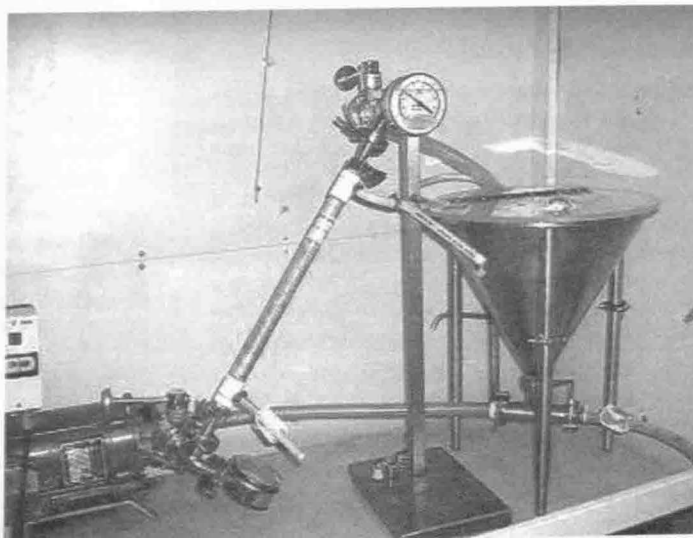
than those of regenerated cellulose membranes. From the above observation, a polyethersulfone membrane with MWCO 10,000 was tested in the Hollow Fiber Ultrafiltration System as a long-term test. As polyethersulfone membrane was not available in a cartridge form in the market, the membrane with the closest molecular structure, polysulfone membrane cartridge, was tested in the Hollow Fiber UF System experiment for the long-term test. The high endotoxin solution with a content of 625 EU/ml was used in this test. Throughout the eight-day experimental period, the membrane rejected endotoxin to less than 0.25 EU/ml without losing its rejecting capability. The flux rate stayed the same at 129 L/m<sup>2</sup>/hr for all the entire period. Therefore, it is strongly recommended that an ultrafiltration method using polysulfone membrane with MWCO 10,000 be used for the primary purification method in producing water for injection.

#### **References:**

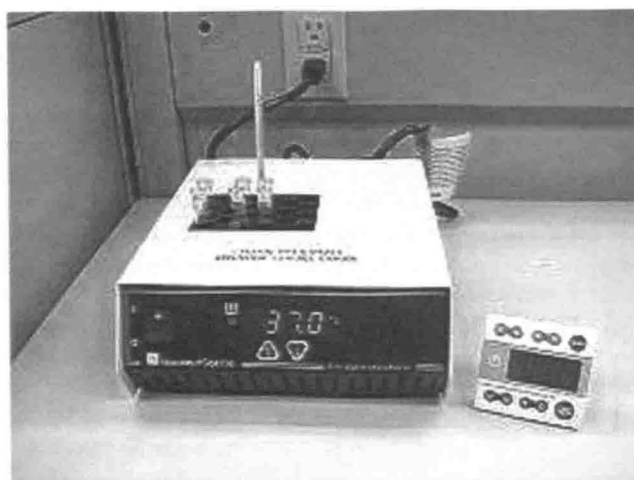
- Cheryan, Munir, "*Ultrafiltration and Microfiltration Handbook*," Technomic Pub. Co. (Lancaster, PA) 1998
- Felice, Antonio, "*Ultrafiltration and endotoxin removal from dialysis fluids*," *Kidney International*, Vol. 43, Suppl. 41 (1993), pp. S-201-S-204
- Grandis, Peter; Szarthmary, Susan "*Method of removing pyrogens from dialysate*" Cont-in-part of U.S Ser. No. 798,432 abandoned.
- Hamilton, R.; Ford, C.; Colton, C.; Cross, R.; Steinmuller, S. and L. Henderson, "*Blood Cleansing by Diafiltration in Uremic Dog and Man*," Vol.XVII Trans. Amer. Soc. Artif. Int. Organs (1971)
- Henderson, L. W. and E. Bean, "*Successful Production of Sterile Pyrogen-Free Electrolyte Solution by Ultrafiltration*," *Kidney Int.* Vol. 14, pg. 522 (1978)
- Stephane, Mabie; Darbouret, Daniel "*Qualification study of a point-of-use ultrafiltration device for the production of pyrogen-free ultra-pure water*," Laboratory Water Division, Millipore S.A. Saint-Quentin-en-Yvelines, France.
- Tutunjian, RS, "*Pyrogen removal by ultrafiltration*," Progress in clinical and biological research, Amicon Corporation, Danvers, MA, 1982.
- USP DI, Vol. III, "*Approved Drug Products and Legal Requirements*," 19<sup>th</sup> ed., United States Pharmacopeial Convention, 1999
- <http://www.acciusa.com/lal/>
- <http://www.millipore.com/publications.nsf/docs/5f5nm>



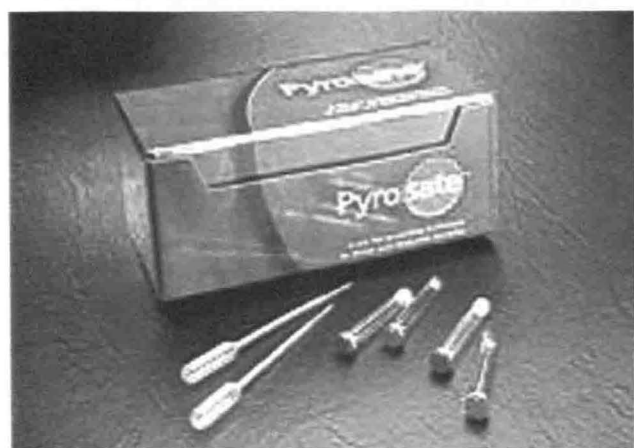
Photograph 1: Gel-Clot Formation a#3DDB1B



Photograph 2: Hollow Fiber Cartridge#3DDB1C



Photograph 3: Incubating a Filtrat#3DDB1D

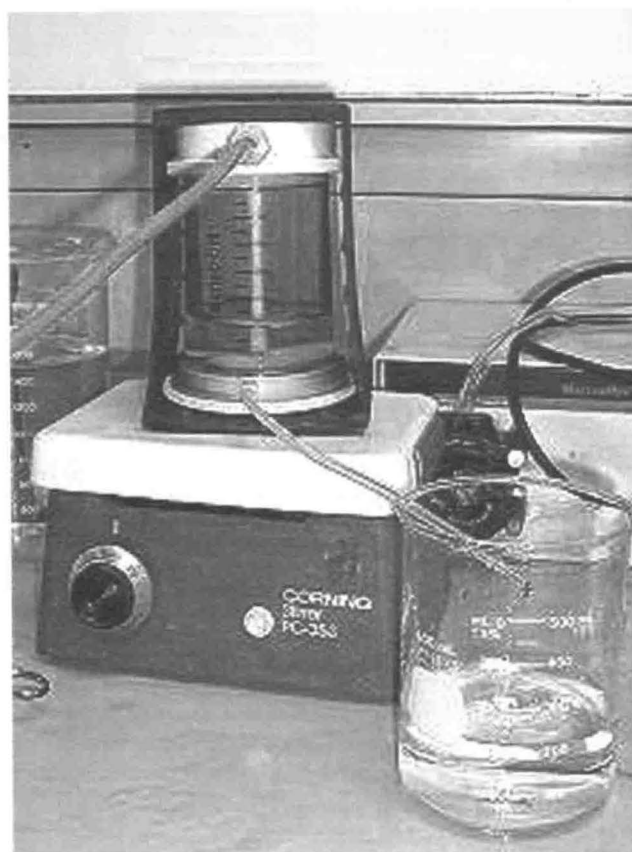


Photograph 4: Pyrosate kits

#### Faculty comments:

Mr. Soe's mentor, Robert Cross, had the following remarks to make about his student's work:

Thet came to see me two years ago searching for an undergraduate research project. We agreed that he would take on a critical part of the work that I was doing under a grant from the Arkansas Biosciences Institute to develop an economical process for preparing the type of high purity water needed in an artificial kidney process called hemodiafiltration. Thet's project was to find the most cost-effective ultrafiltration membrane for use in the process. Thet wrote an excellent proposal and was able to receive an



Photograph 5: Stirred Ultrafiltrat#3DDB1F

undergraduate research scholarship from the Honor's College for his project.

During the course of the project he did outstanding work in formulating the work plan, in carrying out the research, and in writing up the results. In particular, he was able to develop innovative procedures to obtain reproducible results from an analytical test requiring very careful techniques to prevent sample contamination. He was also able to organize the data in such a way to obtain logical conclusions and achieve the project goals.

## DYNAMIC COMPOSITION OF AGENT GRAMMARS

By Kyle Neumeier

Department of Computer Science and Computer Engineering

Faculty Mentor: Craig Thompson

Department of Computer Science and Computer Engineering

### Abstract:

*In the very near future, as pervasive computing takes root, there will be an explosion of everyday objects that are uniquely identifiable and wrapped by a computational layer – effectively bringing the object to life. An important component of this system is the mechanism that will allow humans to interface with the objects. Menu Based Natural Language Interfaces (MBNLI) seem like a good candidate for this job because of the intuitive way in which they allow the user to build commands. However, the MBNLI system will have to scale with the number of objects in the system. This project describes context free grammar modules which are small grammar files that can be composed to form a larger grammar. Grammar modules can then be associated with individual objects, and in this way allow the MBNLI to scale according to the size of the system.*

### 1 Introduction:

#### 1.1 Problem

In a world where electronic devices are shrinking rapidly and the ability to associate unique identifications with objects is becoming commonplace, computing is escaping the desktop and becoming pervasive. The idea behind pervasive computing is that computers are migrating quickly from the familiar desktops, laptops, and PDAs that people use and are being incorporated into everyday objects. These objects – refrigerators, coffeemakers, sprinkler systems, web pages, etc. – will look the same as they do today, but they will be wrapped in a computational process that will give them the ability to communicate with people, the outside world, and even other objects via a programmatic interface and often a wireless connection [1].

In order to scale such a collection of network devices, or more generally a system of *agents* – everyday objects with a computational wrapper – humans will need some sort of control mechanism(s). One possible solution is to associate an interface grammar, or a set of rules describing the object's command set, with each object. The grammar could then be downloaded into a sort of next generation remote control which could in turn be

used by a human to control any object that had such a grammar [2]. In order to accomplish this, the grammars would need to be able to be composed so that larger grammars could be synthesized from smaller ones in such a way that allowed for the ability of grammars to be “plugged in” and allowed for the reuse of common grammars.

Furthermore, the actual mechanism of communication with these agents must be intuitive. It is reasonable to think that an easy, familiar way for humans to communicate with machines is through the same mechanism that we use to communicate with each other, that is, through our natural languages (e.g. English). Therefore, the system of grammars that allow a user to control the agent should allow commands to be issued via natural language.

#### 1.2 Objective

The objective of this project is to define a system for creating multiple grammars (called grammar modules) that can be composed to build a larger grammar. Such a system would provide a natural language interface for a multi-agent architecture.

#### 1.3 Scope

The focus of this thesis is dynamic grammar composition for use with a menu-based natural language system (described below). A smart home RFID application is assumed but the results from this thesis are relevant to a much broader collection of applications that could include asset management, robots, semantic web, and other applications.

#### 1.4 Organization of this Thesis

Chapter 2 covers background information about Menu Based Natural Language Interfaces (MBNLI), the Everything is Alive (EiA) architecture, and grammar composition. Chapter 3 explains an approach to creating a system of composable grammars via context free grammar modules. Chapter 4 discusses the implementation of the system. Chapter 5 states the conclusions and future work.

## 2 Background:

### 2.1 Key Concepts

In order to better explain the problem that a distributed grammar would solve and the method of creating a system of composable grammars, a brief overview of a related concept is provided. A discussion of the Everything is Alive project will clarify the reason that a composable grammar system is desired in the first place. Then an introduction to Menu Based Natural Language Interfaces will show how grammars can be used to build a natural language interface to an object. Next, context free grammars are reviewed and a few terms relating to composable grammars are defined. Finally, attributed grammars will be introduced.

#### 2.1.1 *Everything is Alive Agent System Project*

The Everything is Alive (EiA) project at the University of Arkansas aims to develop a road map of pervasive computing. As an ever increasing number of objects become identifiable by a computer through technologies such as RFID and IPv6, there will be a tremendous explosion of things that we can control via a computer [1]. One goal of the EiA project is to develop an architecture that organizes these objects with a communication wrapper, which can be viewed as a kind of agent, into a system that optimizes usability and scalability [3].

An example of a system of agents at work in the context of the EiA world is a smart sprinkler system. Imagine that a flower bed agent has a sensor device that can measure the amount of water in its soil. When the soil is too dry for its flowers, it sends a message to the sprinkler agent that tells it to turn on. The sprinkler then asks the weather agent on the Internet if it will rain in the next 24 hours. If no rain is likely, then the sprinkler agent turns on and waters the flowers [1]. A second example is a thermostat that knows to turn the air conditioner on when a light turns on<sup>1</sup>. The point is that the objects are the same as before – the sprinkler is still a sprinkler and the thermostat is still a thermostat – but an agent wraps the object, effectively bringing it to life.

Past work in the EiA project has produced an architecture that facilitates agent-to-agent communication [4]. The agent class “wraps” an object and allows it to communicate with other agents by making available methods that send and receive XML messages. When XML messages are received by an agent, they are translated into the underlying object’s native language so that the command can be executed. A few agents have been constructed, including an RFID reader agent that facilitates commands such as “turn on,” “read for 300 ms” and “turn off [5].”

#### 2.1.2 *Menu Based Natural Language Interfaces*

Building natural language interfaces to machines has been a grand challenge problem for almost as long as computers have

existed. The idea is simple; it would be nice if a user could control a computer simply by speaking, as if to another person. The implementation, however, has proved to be extraordinarily difficult. Problems arise because computers cannot interpret connotations, clichés, body language, and idioms that all contribute to our understanding of language. The resulting situation is that the user either *overshoots* the ability of the NLI system by phrasing a command that cannot be understood, or *undershoots* the ability of the underlying system by not using features that are available because the user is not aware of them or not sure how to phrase the command to use them. This mismatch between the user’s phrasing and the NLI system’s capability is known as the habitability problem [6].

Menu Based Natural Language Interfaces (MBNLI) relieve the habitability problem by employing a predictive parser and cascading menus to present the user with a list of next possible choices. When the user selects a phrase from the menu, the parser generates a new list of next possible choices. This process continues until a complete sentence or command is built. Such a system guarantees that any command the user builds is syntactically correct. Furthermore, the user can develop an understanding of the capabilities of the underlying system by exploring the menus [6].

One particularly useful application of MBNLI technology is when it is used as the front end to a relational database. The interface enables users to build queries in English rather than SQL, allowing users who know nothing about SQL or about the database schema (i.e. relation and attribute names) to extract useful information from the database. In the past, NLIs to databases have typically been question and answer systems that allow users to ask natural language questions; however, these systems suffer from the habitability problem in that a user does not know what is able to be asked and exactly how to ask it [7].

The Everything is Alive project has found its own use for MBNLI. The project envisions a world in which many objects in the form of agents can be controlled by humans. Thus, an easy and standard way of issuing commands will be needed. Menu Based Natural Language Interface technology offers a solution to this problem. If all agents had an associated grammar, an MBNLI interface could be generated based on the grammar. The user would then be able to issue syntactically correct commands by building them. Furthermore, the user would know exactly what capabilities the agent has (by virtue of the cascading menus) [8].

#### 2.1.3 *LingoLogic*

LingoLogic is an implementation of an MBNLI system created by Object Services and Consulting, Inc. It was based off of an earlier implementation developed at Texas Instruments in the 1980s called NLMenu. The intent of the project was to develop an MBNLI generator for a relational database. The idea was that if a static grammar and translation for SQL was



developed, MBNLI interfaces could be generated relatively easily by providing the parser with a description of the particular database's schema [7].

LingoLogic consists of a front-end, implemented in Java, and a parser, implemented in C. The parser works like a LISP interpreter in that it is basically a read-eval-print loop. In other words, a command is given to the parser which processes the command immediately. The front end is a GUI that allows users to build commands and queries via a cascading menu. When a user has selected a word or phrase, the front end passes the selected item to the parser via a port, the parser predicts a set of next legal phrases, and sends them back to the front end. This process is repeated until a complete sentence is built. If a target language is specified, the parser can translate the complete sentence into the language. It is then the task of the front end to execute the translation (in whatever sense is appropriate). Because LingoLogic was intended as an interface generator for relational databases, its front end has the ability to execute SQL queries against a database and display the results [7].

#### 2.1.4 Context Free Grammars

LingoLogic uses attributed context free grammars (CFG) to specify the syntax of commands and queries. Context free grammars consist of a finite set of terminals (T), a start symbol (S) which is a member of V, a finite set of nonterminals (V), and a finite set of production rules (P) that represent the recursive definition of a language. The productions take the following form:

Left Hand Side (LHS) → Right Hand Side (RHS)

The LHS is a nonterminal that is being defined. The RHS is a string of terminals and non terminals that represent a way to form the LHS. Formally, the production rules may be defined as follows [9]:

$$\alpha \rightarrow \beta, \text{ where } \alpha \in V, \text{ and } \beta = (T \cup V)^*$$

A very simple version of the English language can be specified using a CFG<sup>2</sup>.

```
S → nounPhrase verbPhrase
nounPhrase → article noun
verbPhrase → verb nounPhrase
article → THE | A
noun → DOG | CAT
verb → EATS | CHASES
```

Example 1: A simple English grammar

This CFG allows the construction of sentences such as "The dog chases a cat" and "The cat eats the dog."

#### 2.1.5 Context Free Grammar Closure Under Union Operation

Central to the idea of composing smaller grammars to create larger ones is the concept of grammar *union*, because a larger grammar can be treated simply a union of smaller ones. It is well known that context free grammars (CFGs) are closed under the union operation. A proof of this fact can be found in most textbooks on formal languages<sup>3</sup>. This result provides a theoretical basis for creating context free grammar modules that are composable into larger grammars [9].

#### 2.1.6 Attributed Context Free Grammars

LingoLogic grammars, however, allow for more expressiveness than would normally be the case with a standard CFG due to the use of attributes. Attributes are extra values attached to the terminals in the form of name-value pairs. The values can then be used in the grammar rules to add constraints to rules. It is these constraints that give LingoLogic its expressive power [10]. For example, suppose that the nonterminals in Example 1 had an extra value called *number* associated with them. The nonterminals might be re-written as follows<sup>4</sup>:

```
noun → DOG[number=singular] |
      DOGS[number=plural] |
      CAT[number=singular] |
      CATS[number=plural]
verb → EATS[number=singular] |
      EAT[number=plural] |
      CHASES[number=singular] |
      CHASE[number=plural]
```

A constraint could then be added to the nounPhrase and verbPhrase rules so these nonterminals adopt the *number* attribute of their terminals<sup>5</sup>.

```
nounPhrase → article noun
[nounPhrase.number = noun.number]
verbPhrase → verb nounPhrase
[verbPhrase.number = noun.number]
```

Finally, a constraint to the top level rule could specify that the number of the noun-phrase and verb-phrase must agree.

```
S → nounPhrase verbPhrase
[nounPhrase.number == verbPhrase.number]
```

Example 2: A simple attributed English grammar

Adding these attributes allow the parser to distinguish between singular and plural nouns and verbs, which means that

sentences such as "Dogs eat cats" are allowed, while sentences such as "Dogs eats cat" are not.

## 2.2 Translations

The LingoLogic parser also has the ability to translate a sentence into a target language [10]. Translations can take many different forms. In the case of relational databases, the natural language queries that are formed via the cascading menus are translated into SQL. The agent system created by the EiA project requires its messages to be XML, so translation rules could be written to build well-formed XML messages. A third idea for a target language is a function call that could then be evaluated by another program. An example of this would be the sentence "Microwave cook for 30 seconds" could be translated to a call to a *microwave* method with arguments *cook* and *30 seconds*. It might look like this:

```
Microwave(cook,30);
```

## 2.3 Advantages of Distributed Grammars

In general, distributed systems have several advantages over stand-alone type systems. The first is that distributed systems can be more robust in that they eliminate single points-of-failure. For example, in a packet switched network, such as the Internet, if a router crashes, packets are still able to reach their destination via another route [11]. In the same way, if a very large grammar file is broken into smaller pieces, these pieces could be stored redundantly in different places, allowing the entire system to function, even if a failure occurs with one piece of the grammar. Secondly, distributed systems can be more efficient because data that is not relevant does not need to be processed. In terms of a distributed grammar, smaller grammar files could be downloaded and composed on the fly, eliminating the need to process rules and lexicons that will not be used. Finally, distributed systems are scalable. This characteristic is important and is one of the driving factors for creating grammar modules in the first place, because it will allow for grammars to be written for agents as they are made. This means there is no need to update a central grammar each time a new agent is added to the system; rather, when a new agent is added, its grammar will be an extension and processed as it is needed.

## 2.4 Related Work

The idea of adding the ability for grammars to be composed of smaller grammars in order to allow for them to be distributed is not new. Because of the obvious benefits of flexibility and robustness that would be gained from this capability, distributing grammar files was a central concept in OBJS' quest to enable agents to be controlled via an MBNLI on the Web.

### 2.4.1 AgentGram

AgentGram was a prototype of a MBNLI system that was developed from 1998-1999. Unlike LingoLogic, it did not parse

context-free languages; instead it handled only a simpler tree-grammar and did not handle translations. AgentGram did, however, have the ability to load grammar files on-the-fly, thus allowing grammars to be chained together. When a terminal or nonterminal that was not specified in the current grammar file was reached, a URL pointed to the grammar file that could complete the rule. This URL was then followed, the grammar downloaded and processed, and the parse continued seamlessly [12]. For example:

```
<item name = "List the">
  <item name = "hotels">
    <item name = "in L.A.">
      <item name = "[THEN]"
URL = "http://...hotels.xml">
  <item name = "in Washington D.C.">
    <item name = "[THEN]"
URL = "http://...hotels.xml">
```

The file *hotels.xml* would then have some information specific to hotels. It might look something like this:

```
<item name = "where the hotel name is">
  <item name = "Best Western">
  <item name = "Clarion">
<item name = "where the hotel costs">
  <item name = "less than $70 per night">
  <item name = "between $71 and $100 per night">
  <item name = "more than $101 per night">
```

### Example 3: An AgentGram XML grammar

This example should clarify the advantages of a system of distributed grammars. The MBNLI system has to load only relevant grammar rules when it requires them instead of loading all rules that it might ever use. Furthermore, this distributed type of system is easier to scale and maintain because, when the system changes, only relevant files must be updated. For example, if a new hotel were built in L.A. only the hotel grammar would need to be modified.

### 2.4.2 Patent Application

The idea for a system of distributed grammars was conceived for the LingoLogic MBNLI system as well, though it was not implemented. An OBJS patent application [13] describes a method of "chaining" grammars together by encapsulating the grammars (productions, terminals, etc.) into grammar descriptors. Besides containing the grammar, the descriptors would include other information as well, such as a set of pointers to other

descriptors that the grammar references. The parser can either process the references actively by recursively processing the descriptors that the higher level descriptors point to until there are no more links, or lazily by processing a descriptor only when needed.

### 3 Approach:

A multi-agent system will contain many agents of various kinds – various human roles as well as vehicles, equipment, pets, sensors, and even passive things like pictures. Assume that each agent can have an associated grammar. In order to scale a multi-agent system, agents can come and go so it must be possible to add or remove grammars dynamically. This means that grammars will have to be loaded on the fly and that grammars of the system will need to be able to be composed dynamically. There are various ways to do this. The LingoLogic patent application [13] describes one way involving grammar descriptors and chaining grammars together. In spite of a limitation of the current implementation of LingoLogic (that all nonterminals for a grammar must be known before any rules are specified), we can still simulate breaking up the grammar files for a system into grammar modules.

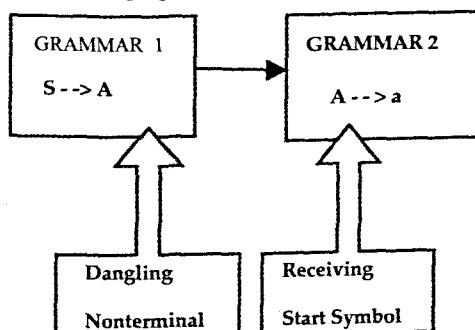
#### 3.1 Context Free Grammar Modules

In their most basic form, CFG Modules are simply CFGs that have been broken into semantic groups. Each group is a CFG in its own right, meaning that it has a start symbol, a set of nonterminals, a set of terminals, and a set of productions. The difference is that a nonterminal in the RHS of a higher level grammar will link to the start symbol of a lower-level CFG module, enabling the chaining of several smaller grammars into a larger one.

##### 3.1.1 Terminology

For the sake of clarity, two terms will be defined. A *dangling nonterminal* is the RHS nonterminal in a higher level grammar that links it to the start symbol of a lower level grammar. A *receiving start-symbol* is the LHS nonterminal of a lower level grammar to which a dangling nonterminals points.

Figure 1: A dangling nonterminal and a receiving start symbol



#### 3.1.2 Types of CFG modules

When using CFG modules as a way to distribute agent MBNLI grammars, there are two possible reasons for creating a certain module. The first is that a “plug-in” ability is desired. This is achieved when one dangling nonterminal points to several receiving start symbols. The second is when module reuse is desired. This is achieved when several dangling nonterminals point to one receiving start symbol. Although these two types of CFG modules are not mutually exclusive, both types will be examined distinctly.

##### One-to-many modules

In order to scale the NLI system to many agents, it will need to be easy to add or “plug in” an agent to the system. CFG modules provide this feature naturally. In this case, a higher level dangling nonterminal points to the receiving start symbols of several agent grammars.

```

S      - > AGENT

AGENT - > agent A

AGENT - > agent B
  
```

##### Many-to-one modules

Sometimes several agents will share similar features. In this case, an ability to reuse grammar files is desired. CFG modules support this scenario by allowing several dangling nonterminals to point to the same receiving start symbol.

```

AGENT _ A      -> FEATURE _ 1

AGENT _ B      -> FEATURE _ 1

FEATURE _ 1    -> foo bar
  
```

#### 3.2 Smart House: An Example

The idea behind CFG Grammar modules will become clear with an example. A smart house example will be used. Imagine a house full of normal objects that each have a wrapper that allows them to be identified as unique objects by a computer and can be controlled via devices that download the object’s grammar. The complete grammar of the house might look like this:

```

S -> microwave MICROWAVECOMMAND
    | oven OVENCOMMAND
    | thermostat THRMSTCOMMAND
  
```

MICROWAVECOMMAND -> turn on

| turn off  
| time for TIME  
| cook for TIME

OVENCOMMAND -> turn on

| turn off  
| time for TIME  
| cook for TIME  
| preheat until temp is  
PREHEAT-TEMP

THRMSTCOMMAND -> turn on heat in ROOMS

| turn on heat in ROOMS until  
temp is ROOM-TEMP  
| turn on air in ROOMS  
| turn on air in ROOMS until  
temp is ROOM-TEMP  
| turn off heat in ROOMS  
| turn off heat in ROOMS until  
temp is ROOM-TEMP  
| turn off air in ROOMS  
| turn off air in ROOMS until  
temp is ROOM-TEMP

TIME -> 30s | 1m | 2m | 3m | 4m | 10m

PREHEAT -> TEMP -> 300 | 325 | 350 | 374 | 400 |  
450

ROOMS -> ROOM

| ROOM and ROOMS

ROOM -> living room | dining room | master  
bedroom | kitchen

ROOM -> TEMP -> 60 | 65 | 70 | 75 | 80

Example 4: The smart house grammar

### 3.3 Forming CFG Modules

This grammar can be divided into semantic groups, and the groups can be represented as files as follows:

\*\*\*\*\*

MAIN GRAMMAR

S -> DEVICE-AND-COMMAND

\*\*\*\*\*

MICROWAVE GRAMMAR

DEVICE-AND-COMMAND -> microwave  
MICROWAVECOMMAND

MICROWAVECOMMAND -> POWER | TIMER |  
COOK

\*\*\*\*\*

OVEN GRAMMAR

DEVICE-AND-COMMAND -> oven  
OVENCOMMAND

OVENCOMMAND -> POWER | TIMER | COOK |  
PREHEAT

\*\*\*\*\*

THERMOSTAT GRAMMAR

DEVICE-AND-COMMAND -> thermostat  
THRMSTCOMMAND

THRMSTCOMMAND -> POWER heat in ROOMS

| POWER air in ROOMS

| POWER heat in ROOMS

until temp is ROOM-TEMP

| POWER air in ROOMS

until temp is ROOM-TEMP

| display\_temperature in

ROOM

\*\*\*\*\*

POWER GRAMMAR

POWER -> turn on | turn off

\*\*\*\*\*

TIMER GRAMMAR

TIMER -> time for TIME

\*\*\*\*\*

COOK GRAMMAR

COOK -> cook | cook for TIME

\*\*\*\*\*

TIME GRAMMAR

TIME -> 30s | 1m | 2m | 3m | 4m | 10m

\*\*\*\*\*

PREHEAT GRAMMAR

PREHEAT -> preheat until temp is PREHEAT-TEMP

PREHEAT-TEMP -> 300 | 325 | 350 | 375 | 400 | 450

\*\*\*\*\*

#### ROOMS GRAMMAR

ROOMS -> ROOM and ROOMS

ROOM -> living room | dining room | master  
bedroom | kitchen

\*\*\*\*\*

#### ROOM-TEMP GRAMMAR

ROOM-TEMP -> 60 | 65 | 70 | 75 | 80

\*\*\*\*\*

#### Example 5: Smart house grammar modules

In this set of example grammar modules, it is easy to see that both the one-to-many and many-to-one module paradigms are used. The one-to-many type modules are used to add new devices to the system. This is possible because the dangling nonterminal in the highest level module is *DEVICE-AND-COMMAND*; adding a new device to the system relatively simple: just make the receiving start symbol of the new device grammar *DEVICE-AND-COMMAND*. Similarly, two modules sharing a module, utilizing the many-to-one feature of CFG modules, is also exemplified in the smart house set of grammar modules. The microwave grammar and the oven grammar both share the cook grammar module. They do this by have a dangling nonterminal called *cook*. The cook module's receiving start symbol is also named *cook*.

### 3.4 Limitations of Grammar Modules

Although the simple conversion of a CFG to a collection of CFG modules addresses several issues that are important to scaling a multi-agent system, it has several inherent limitations. One of the major limitations of grammar modules is that all the modules are global because there is no inherent scoping mechanism. This limitation creates two problems. The first is that there may be situations in which certain users should not be able to control certain agents. The second problem is that certain modules cannot be reused despite the fact that they share similar features. A third problem (unrelated to the scoping problems) with CFG modules is that many times a large (or even infinite) number of nonterminals should be able to be used, but due to the fact that it is difficult to specify ranges of values in CFGs, presenting the user with a list of all possible values from which to choose is not practical.

#### 3.4.1 Security

In certain situations, certain users should not have access to every feature of an entire system. This concept is familiar in the world of operating systems, in which only administrators can control certain programs. Similarly, perhaps a parent does not want a child to be able to control the *power* feature of a thermostat

agent but would like him to be able to issue the *display\_temperature* command. Because all grammar rules are global, a new thermostat grammar module would have to be written that excluded the thermostat rules that began with the nonterminal *power*. This new thermostat grammar module would have to be loaded instead of the other one when the child is using the thermostat. A better solution might be to add a way to scope grammar rules so that the *power* rules are not accessible to the child.

#### 3.4.2 Reusability of many-to-one modules

In a related problem to that of the security issue, CFG modules can only be reused when exactly the same feature set is desired. For example, a new grammar module that allows a user to control lamps is created such that it has the following grammar:

\*\*\*\*\*

DEVICE-AND-COMMAND -> POWER lamp in  
ROOMS

\*\*\*\*\*

This lamp grammar would control lamps in the exact same rooms in which the thermostat can control the temperature. Suppose, however, that there is no lamp in the kitchen. This means that a new *rooms* CFG module must be created that is the same as the old *rooms* module but does not include the kitchen as an available terminal. This inability to scope causes the need for a new grammar module to be created despite the fact that one that is very similar already exists. In a large scale system, the inflexibility caused by the lack of a scoping mechanism might undermine the benefits gained from the ability to reuse CFG modules.

#### 3.4.3 Inability to Specify a Value from a Range of Terminals

Many times the need to be able to specify a value from a range is required. This limitation is illustrated in the smart house example: the *time* module allows only a small handful of values.

One way to solve this problem would be to list every possible value as a terminal. This method, however, would also be cumbersome not only from the grammar writer's point of view, but also from the user's point of view when he has to pick a value from a very large menu. Another possible solution would be to write another grammar for the time module such as the following:

TIME -> MIN SEC

MIN -> DIGIT DIGIT min

SEC -> DIGIT DIGIT sec

DIGIT -> 0|1|2|3|4|5|6|7|8|9

The problem is that this grammar allows values such as 2 min 88 sec, which is not correct.

### 3.5 Using Attributes to Scope CFG Modules

In Section 2.1.6, the idea of adding attributes to CFGs in order to extend the expressiveness of them is discussed. Attributes, which are essentially trees consisting of name-value pairs in the context of LingoLogic, can be used in order to add a sort of scoping mechanism to CFG modules.

#### 3.5.1 Using Attributes to Solve the Security Problem

In the previous security example, a parent would like their child only to be able to use the *display-temperature* command in the thermostat module. In order to use attributes to solve this problem, first assume that a global attribute called *user* was added to the system. The thermostat module could be re-written as follows:

```
*****
THERMOSTAT GRAMMAR

DEVICE-AND-COMMAND - > thermostat
THERMSTCOMMAND

[global.user ∈ THERMSTCOMMAND.users]

THERMSTCOMMAND - > POWER heat in ROOMS
[users = {parent}]

| POWER air in ROOMS [users = {parent}]

| display_temperature in ROOM

[users = {parent, child}]
*****
```

Example 6: Using attributes to add security to the Thermostat grammar

Rewriting the thermostat grammar in this way, only allows the parent to access the *power* commands because the *DEVICE-AND-COMMAND* rule is only able to be used when the *user* attribute of a global attribute tree is a member of the *users* attribute tree of *THERMSTCOMMAND*.

#### 3.5.2 Using Attributes to Increase Reusability of Modules

As mentioned above, a module can be reused only if more than one higher-level modules share the *exact* same feature set. The previously used example was a *lamp* grammar that shares many, but not all, of the same rooms as the *thermostat* grammar. The solution, to create a new *rooms* grammar that contains a subset of the rooms listed in the original grammar used by the thermostat, is not scalable because it would create a multitude of very similar grammars in a large scale system.

Attributes could be applied in a similar way to the solution to the security issue to solve this problem. If an *objsInRoom* attribute were added to each room in the *rooms* grammar, then the set of objects available to be controlled in each room could be specified.

\*\*\*\*\*

#### ROOMS GRAMMAR

ROOMS -> ROOM and ROOMS

[ROOMS.objsInRoom = ROOM.objectsInRoom]

ROOM - > living room [objsInRoom = {thermostat, lamp}]

| dining room [objsInRoom = {thermostat, lamp}]

| master bedroom [objsInRoom = {thermostat, lamp}]

| kitchen [objsInRoom = {lamp}]

\*\*\*\*\*

The thermostat and lamp grammar could be rewritten to test for membership in each of the rooms.

\*\*\*\*\*

#### THERMOSTAT GRAMMAR

DEVICE-AND-COMMAND - > thermostat  
THERMSTCOMMAND

THERMSTCOMMAND - > POWER heat in ROOMS

[thermostat " ROOMS.objsInRoom]

| POWER air in ROOMS

[thermostat " ROOMS.objsInRoom]

| POWER heat in ROOMS until temp is ROOM-TEMP

[thermostat " ROOMS.objsInRoom]

| POWER air in ROOMS until temp is ROOM-TEMP

[thermostat " ROOMS.objsInRoom]

| display\_temperature in ROOM

[thermostat " ROOMS.objsInRoom]

\*\*\*\*\*

#### LAMP GRAMMAR

DEVICE-AND-COMMAND - > POWER lamp in ROOMS

[lamp " ROOMS.objsInRoom]

\*\*\*\*\*

Example 7: Using attributes to increase the reusability of the Room grammar

In this way, attributes can make CFG modules reusable, even if only a subset of its rules or terminals is to be used.

### 3.6 Using Experts to Specify a Value from a Range of Terminals

It is difficult to specify ranges of values from which to choose a value using CFGs. One solution is to let terminals be either values (as they currently are) or function calls that return a value. The function could then have some logic that would allow a user to choose easily a value from a range. For example, the *time* grammar could be changed so that instead of the CFG rules used to specify a time, a function is called that executes code that allows a user to specify the time.

```
*****
TIME GRAMMAR

TIME -> timer()
*****

TIMER FUNCTION

/* Note that this function is for specifying a timer
time, such as "count for 4 hours and 20 minutes." It
is not for specifying a time of day like "4:00 PM". */

1 INT hour, min;
2 PRINT "ENTER HOURS";
3 GET hour;
4 IF (hour < 0) {PRINT "HOUR MUST BE >=0";
GOTO 2}
5 PRINT "ENTER MINUTES";
6 GET min;
7 IF (min > 60) { PRINT "MIN MUST BE < 60;
GOTO 5}
8 IF (min < 0) { PRINT "MIN MUST BE >= 0; GOTO 5}
9 RETURN hour + ":" + min;
*****
```

Example 8: Using an expert to specify a range of timer times

The functions that return values are called *experts* because they are an "expert" at knowing a particular field. It can be seen that experts can help a user choose a value from all kinds of ranges including integers, currency, and time of day. If the logic in the expert were more complicated than the simple *timer* function above, a graphical user interface could be used to specify the value. A useful example might be that the expert produces a color chart and allows the user to visually select a color [14].

## 4 Implementation:

### 4.1 LingoLogic Grammar files

The smart house grammar was implemented using the LingoLogic Toolkit. LingoLogic consists of a predictive parser and an interface. The interface lets the user choose a phrase from a set of choices, each representing continuations of the sentence, and then sends the phrase back to the parser which returns a set of next possible choices. This process is continued until a complete sentence is built. The LingoLogic parser is controlled via Lisp-like commands. A LingoLogic grammar has four parts: the parser initialization, category declaration, lexicon definitions, and rule definitions [10].

#### 4.1.1 Parser initiation

The first step in writing a LingoLogic grammar is to initialize the parser. This involves the following statement, which instantiates a parser object and creates a pointer to the newly created parser object. The parser initialization statement has the following syntax, where *parser1* is the pointer to the parser:

```
(setq parser1 (create-parser))
```

#### 4.1.2 Parser Categories

Secondly, the parser categories are defined through a call to the function *set-parser-categories*. Parser categories are the set of nonterminals that will be used in the grammar. In other words, no nonterminal may be used in the grammar rule unless it was declared to the parser. LingoLogic requires that all parser categories be declared before the first rule is defined. The function *set-parser-categories* may, however, be called multiple times, as long as the final time that it is called is before the final grammar rule is declared. The final set of parser categories is the union of the parser categories declared in each call to *set-parser-categories*. The function *set-parser-categories* has the following syntax, where *parser1* is a pointer to the parser.

```
(set-parser-categories parser1 (<category symbols>))
```

#### 4.1.3 Lexicon

The lexicon is the set of terminals used in a certain grammar. Terminals are defined using a call to the function *defword*, with the following syntax, where *word1* is the name of the word being defined, *parser1* is a pointer to the parser, *<attribute-tree>* is a tree of attributes associated with the word, *<menu>* is the menu on which the word will appear, *<print-info>* is the string that will appear on the menu, and *[expert-info]* is an optional function that may be run to aid the user in specifying a value [15]:

```
(DEFWORD word1 parser1 <attribute-tree>
<menu> <print-info> [<expert-info>])
```



The most complex aspect of this definition is the *<attribute-tree>*. As previously mentioned in Section 2.1.6, LingoLogic has the ability to parse attributed grammars. Attributes are extra values attached to terminals that add semantics not imparted by the syntax of the grammar [16]. In the context of LingoLogic, the attributes are defined as name-value pairs or name-value lists that form attribute trees. Constraints can then be added in the form of rules that use the trees to add expressive power. One attribute, *:cat*, short for category, is required by the parser, because it associates the word with a parser category. The lexicon for the attribute English grammar of Example 2, expressed using calls to *defrule* would look as follows:

```
(defword dog p1 (((:cat noun)(number sing)) nouns
"dog"))

(defword dogs p1 (((:cat noun)(number plural)) nouns
"dogs"))

(defword cat p1 (((:cat noun)(number sing)) nouns
"cat"))

(defword cats p1 (((:cat noun)(number plural)) nouns
"cats"))

(defword chases p1 (((:cat verb)(number sing)) verbs
"chases"))

(defword chase p1 (((:cat verb)(number plural)) verbs
"chase"))

(defword eats p1 (((:cat verb)(number sing)) verbs
"eats"))

(defword eat p1 (((:cat verb)(number plural)) verbs
"eat"))
```

#### 4.1.4 Grammar Rules

The most complex portion of a LingoLogic grammar file is the set of grammar rules. Like all other LingoLogic parser commands, the grammar rules are written using a LISP-like syntax<sup>6</sup>. The rules are defined using the following syntax, where *rule1* is the name of the rule, *parser1* is the pointer to the parser, *<term1>* is the LHS, *<term2>...<term3>* is the RHS that define *<term1>*, and [*<constraint>*] are optional constraints that may be added due to the ability of the parser to handle attributes [10].

```
(defrule rule1 parser1 (<term1> —> <term2>
<term3>...)

[<constraint> ...])
```

In Example 1, a simple English language grammar is introduced. If this grammar were written for LingoLogic, the top level rule would be defined as follows:

```
(defrule rule1 parser1 (S —> nounPhrase verbPhrase))
```

Example 2 augments the simple grammar with attributes. Grammar rules can then use these attributes in constraints that

add expressive power. For example, a constraint which requires the number of the *nounPhrase* and *verbPhrase* to agree can be added to the rule above.

```
(defrule rule1 parser1 (S —> nounPhrase VerbPhrase)

((nounPhrase number) = (verbPhrase number)))
```

#### 4.2 LingoLogic CFG Modules

The first step in implementing CFG modules in LingoLogic is to write the grammar rules and break them into conceptual groups, as in Section 3.3. Then the grammar rules must be translated to the LISP-syntax of the LingoLogic parser command set. Next, the parser category declarations have to be separated into their own files. Finally, a parser initialization file must be created.

##### 4.2.1 Writing LingoLogic Grammar Files

The translation of the CFG production rules into LingoLogic grammar functions consists of writing a *defrule* function for each definition of every nonterminal in the grammar. In other words, for each nonterminal  $\alpha$  in grammar  $G$ , a *defrule* function must be written for every  $\beta$ , where  $\alpha \rightarrow \beta_1 | \beta_2 | \dots | \beta_n$ . For example, the microwave module has a production rule that looks like this:

```
MICROWAVECOMMAND -> POWER | TIMER |
COOK
```

The translation of this rule into a LingoLogic grammar rule involves writing three *defrule* functions.

```
(defrule rule1 parser1 (MICROWAVECOMMAND
—> POWER))

(defrule rule2 parser1 (MICROWAVECOMMAND
—> TIMER))

(defrule rule3 parser1 (MICROWAVECOMMAND
—> COOK))
```

Furthermore, a *defword* function must be written for every terminal in the grammar. The *:cat* attribute of the word must be the nonterminal which is defined by that word. For example, the room module has a production rule that looks like this:

```
ROOM -> living room | dining room | master
bedroom | kitchen
```

In order to write the LingoLogic translation of this rule, a *defword* function must be written for each room.

```
(defword livingroom parser1 (((:cat room)) rooms
"Living Room"))

(defword diningroom parser1 (((:cat room)) rooms
"Dining Room"))

(defword mbedroom parser1 (((:cat room)) rooms
"Master Bedroom"))
```

```
(defword kitchen parser1 (((:cat room)) rooms
"Kitchen"))
```

One important detail of LingoLogic *defrule* function is that the RHS of the productions may consist only of pointers to parser categories, which implies that the RHS of a production rule may *not* contain a pointer to a *defword* function. This means that terminals are included in grammar rules indirectly by means of the parser category (:cat) attribute of the *defword* function.

A final aspect of translating a CFG into LingoLogic grammar functions is the rule namespace. Each *defrule* and *defword* function is global, so care must be taken not to give two rules or two words the same name.

#### 4.2.2 Creating the Parser Category Files

The LingoLogic parser requires that all parser categories be declared before the first grammar rule is defined. This requirement is a result of the way the parser handles the construction of some internal data structures. Therefore, for each grammar module, a parser category file must be created. This file consists only of a call to *set-parser-categories* (see Section 4.1.2) and declares parser categories on the RHS of every grammar rule in the file. In this way, the parser can first execute all of the *set-parser-categories* functions by reading all of the parser category files before it reads any of the *defrule* or *defword* functions.

#### 4.2.3 Parser Initialization File

The final file that must be created is a parser initialization file, which has three parts. The first part is to create the parser with a command such as the following:

```
(setq parser1 (create-parser))
```

The second step is to load all of the parser category files.

```
(load "main.pc") ; ; load main parser categories.
```

```
(load "microwave.pc") ; ; load microwave parser
categories
```

```
(load "oven.pc") ; ; load oven parser categories
```

...The third step is to load all of the grammar files.

```
(load "main.gnl") ; ; load the main grammar
```

```
(load "microwave.gnl") ; ; load the microwave
grammar
```

```
(load "oven.gnl") ; ; load the oven grammar
```

Finally, a batch file is created that starts the parser and initializes it with the initialization file and starts the front end.

## 5 Conclusions:

### 5.1 Significance

As more and more objects are added to the Everything is Alive agent system project, a standardized means to control them will be needed. LingoLogic, due to its intuitive, guided approach to issuing commands is a good candidate for the mechanism to control agents. In order for LingoLogic to be useful, however, the agent grammars must be able to scale. Context Free Grammar modules provide a means to create scalable agent grammars that can be composed together at run time to create a menu based natural language interface to the agent system.

### 5.2 Future Work

One area of improvement for this project involves the binding time of the parser. Because the parser requires all parser categories to be declared before the first grammar rule is defined, grammar rules that involve new categories cannot be added after the parser has been initialized. This means that grammars cannot be added on the fly. It would be better if the parser were improved to accept new parser categories even after grammar rules have been declared. Then, grammars could be truly dynamic, in that new modules could be added on the fly.

Along with the ability to add new grammars on the fly, it would be useful if grammar descriptors, as described in [13] were developed. Grammar descriptors would provide a better encapsulation model for LingoLogic grammars, because a descriptor would be meta information at the top of the grammar file that describes how it is chained to other grammars.

Finally, another area of improvement to this project would be to add the ability to detect which agents are connected to the agent system, allow the user to select which agents he would like to communicate with, then pull the selected agents' grammar modules from a database. Such a system would simulate how a soft controller might work in the future.

## 6 References:

- [1] C. Thompson, "Everything is Alive," Architectural Perspective Column, *IEEE Internet Computing*, Jan-Feb 2004. <http://csce.uark.edu/~cwt/DOCS/2004-01-PAPER-IEEE-Internet-Computing-Everything-is-Alive.pdf>
- [2] C. Thompson, "Smart Devices and Soft Controllers," Architectural Perspective Column, *IEEE Internet Computing*, Jan-Feb 2005. <http://csce.uark.edu/~cwt/DOCS/2005-01-PAPER-IEEE-Internet-Computing-Smart-Objects-and-Soft-Controllers.pdf>
- [3] Vu, Minh, Craig Thompson, "E2 Agent Plugin Architecture," *IEEE International Conference on Integration of Knowledge Intensive Multi-Agent Systems (KIMAS-05)*, Waltham, MA, April 18-21, 2005. <http://csce.uark.edu/~cwt/DOCS/2005-01-PAPER-KIMAS-05-Vu-Thompson-E2-agent-plugin-architecture.doc>
- [4] J. Robertson, C. Thompson, "Everything is Alive Agent Architecture," *IEEE International Conference on Integration of Knowledge Intensive Multi-Agent Systems (KIMAS-05)*, Waltham, MA, April 18-21, 2005. <http://csce.uark.edu/%7Ecwt/DOCS/2005-01-PAPER-KIMAS-05-Robertson-Thompson-Everything-is-Alive-Agent-System.doc>
- [5] J. Hoag, C. Thompson, "RFID Agent Middleware Architecture,"

Conference on Applied Research in Information Technology, Conway, AR, March 3, 2006. <http://csce.uark.edu/~cwt/COURSES/2006-01—CSCE-490-590—RFID-Agent-Middleware/DOCS/2005-12—PAPER—ALAR—RFID-Agent-Middleware—Hoag-Thompson—long.doc>

[6] C. Thompson, P. Pazandak, "Introduction to Menu-based Natural Language Interfaces," Technical Memo, Object Services and Consulting Inc., 2000, <http://www.objs.com/agility/tech-reports/0101-MBNLI.doc>

[7] C. Thompson, G. Hansen, "NLI Query Interface," Object Services and Consulting Inc. September 1998, <http://www.objs.com/OSA/NLI-Query-Service.html>

[8] C. Thompson, "Talk to your Semantic Web," Architectural Perspective Column, *IEEE Internet Computing*, Nov-Dec 2005.

[9] J. Hopcroft, R. Motwani, J. Ullman, *Introduction to Automata Theory, Languages, and Computation*, Addison-Wesley, 2001.

[10] G. Lystad and R. Roberts, Chapter 5 in "LingoLogic Manual," Object Services and Consulting Inc., 2000.

[11] J. Kurose, K. Ross, *Computer Networking*, United States of America, Pearson Education, 2005

[12] P. Pazandak, C. Thompson, "AgentGram: Natural Language Interface for Agents," Project Summary, Object Services and Consulting Inc. June 2002, <http://www.objs.com/agility/final/AgentGram/AGENTGRAM-PROJECT-SUMMARY.html>

[13] P. Pazandak, C. Thompson, "Guided Natural Language Interface System and Method," Patent Application, August 2000.

[14] G. Lystad and R. Roberts, Chapter 9 in "LingoLogic Manual," Object Services and Consulting Inc., 2000.

[15] G. Lystad and R. Roberts, Chapter 4 in "LingoLogic Manual," Object Services and Consulting, 2000.

[16] "Attribute Grammar," Wikipedia.com, [http://en.wikipedia.org/wiki/Attribute\\_grammar](http://en.wikipedia.org/wiki/Attribute_grammar)

### Endnotes:

<sup>1</sup> Although these examples may seem trite, imagine the benefits of seamless object interaction in a hospital or an airport.

<sup>2</sup> Please note the convention that nonterminals are lowercase and terminals are upper-case

<sup>3</sup> See Ullman, pg. 284 for one example of this proof.

<sup>4</sup> By convention, brackets indicate an attribute specification.

<sup>5</sup> Note the convention that the dot indicates selection, the equal sign indicates an assignment and two equal signs indicate a test for equality.

<sup>6</sup> The LISP-syntax of the parser command set is an artifact of the fact that LingoLogic's predecessor, NLMenu, was developed at Texas Instruments on LISP Machines in the 1980s. Although LISP is not as popular today, its advantages should not be overlooked: LISP's ability to manipulate symbols accounts for the resemblance of grammar rule definitions to the traditional CFG arrow notation in the formal languages literature.

### Faculty comments:

Dr. Craig Thompson, Mr. Neumeier's mentor, made the following comments about his student's work:

First, a few words about Kyle – as it happens, this afternoon, Kyle is receiving an award as the Top Undergraduate Senior in Computer Science from the CSCE Department in the annual College of Engineering end of year meeting. This follows on the heels of Kyle providing technical support for a Walton School of Business entrepreneurial team that placed in the top ten out of 100 in the U San Francisco Business Plan Competition held over Spring Break, beating out MIT, Harvard and other top schools, and, more recently, also placing second in the Arkansas Governor's Cup Business Plan Competition and winning the Technology Award (based on the technology described below). Finally, Kyle has served

informally as TA in my graduate class on Natural Language Interfaces this semester.

Now, about Kyle's work: Over the last year, Kyle received two undergraduate research grants to work with me on menu-based natural language on the topic of grammar composition. Imagine a collection of software systems (called agents) that can communicate with each other. Each one might represent and control a different device or thing in a home, office, or anywhere (cars, robots, sensors, thermostats, pictures, Ö). Imagine each has an RFID tag (which is like a barcode that can be read at a distance). Now imagine when a person points a next-generation truly "universal remote" at these things, a grammar for controlling that thing is automatically downloaded into the universal remote. This would allow the person to communicate or control things in an unprecedented manner. Add to this that the grammar is in a sort of domain-restricted English and uses menus so a user always knows what they can talk to things about, even if they have never seen that particular thing or kind of thing before. Kyle's thesis is focused on the part of this vision that involves downloading the grammars and composing grammars so the user of the universal remote can control multiple things at once. This is a significant step towards pervasive computing where computing is embedded in the world around us, not just on our desktops.

## A STUDY OF THE IMPACT OF PROGNOSTIC ERRORS ON SYSTEM PERFORMANCE

By Mauricio Carrasco  
Department of Industrial Engineering

Faculty Mentor: C. Richard Cassady, Ph.D., P.E.  
Department of Industrial Engineering

### Abstract:

*Condition-based maintenance is growing in popularity as a means of improving equipment maintenance efficiency. Whether it be the maintenance of an airplane, a computer system, or any type of physical system, the prognostic tools associated with condition-based maintenance are subject to statistical error. These errors can lead to unnecessary preventive maintenance due to underestimation of system remaining life and unnecessary system failures due to overestimation of system remaining life. What is not clear is if these statistical errors outweigh the benefits of a condition-based maintenance policy. This study attempts to address this concern through the evaluation and comparison of three maintenance policies for a simple system. The maintenance policies are run-to-failure, scheduled preventive maintenance and condition-based maintenance. A discrete-event simulation model is used to estimate the average time between successful missions for the system under each of these policies. An extensive set of numerical experiments is used to analyze system performance under a wide variety of operating conditions. The results suggest that condition-based maintenance can improve system performance as much as 10% to 15% beyond that achieved using scheduled preventive maintenance. However, the results also suggest that moderate statistical error can render condition-based maintenance inferior to scheduled maintenance and severe statistical error can render condition-based maintenance inferior to run-to-failure. In addition to the results obtained by this study, the methodology used herein can aid maintenance managers in moving from a scheduled maintenance philosophy to a just-in-time maintenance philosophy; thereby increasing the availability of affected systems. Increasing the availability of any system is given considerable importance especially by industries that serve people. For example, in the airline and health industries the availability of a system is vital since any associated down time results in large profit losses and customer dissatisfaction. Overall, the method presented herein can help any kind of industry in developing a way for assessing their maintenance policies which could help them improve the availability of their systems in the future.*

### Introduction:

The use of prognostics and condition-based maintenance has recently received an increased amount of interest from many industries. These methods use some physical assessment of a system to predict its remaining life and take maintenance action if appropriate. Ideally, such an action will take place instantaneously before failure so that no failures occur and no system uptime is lost unnecessarily. However, the challenge associated with prognostics is developing a system assessment mechanism that is both economically feasible and statistically valid as a means of predicting the remaining system life. Herein, the focus is on the second aspect of this challenge – statistical errors. The main objective of this research is to demonstrate a potential method for evaluating the impact of prognostic errors on system performance. To achieve this objective, a discrete-event simulation model is used to assess the performance of a system under three maintenance policies: (1) run-to-failure maintenance, (2) scheduled preventive maintenance, (3) condition-based maintenance. Various levels of prognostic error, including the ideal case in which prognostics are perfect, are modeled. The results of this experimentation are used to address three questions: (1) How much can perfect prognostics improve system performance beyond scheduled preventive maintenance? (2) How bad do prognostics have to be to make things worse than scheduled preventive maintenance? (3) How bad do prognostics have to be to make things worse than run-to-failure maintenance?

### Model Development:

The goal of this research is to demonstrate a potential method for evaluating the impact of prognostic errors on system performance. To achieve this objective, a discrete-event simulation model was built to assess the performance of a simple system under three maintenance policies. The remainder of this section introduces the system being considered, describes the three maintenance policies, and explains the logic and assumptions behind the simulation model.

Consider a system that can be represented by a single, "black box" component. A new copy of this component has a

Weibull time to failure  $X$  with cumulative distribution function  $F(x)$  having shape parameter  $\beta > 1$  and scale parameter  $\eta > 0$ ,

$$\text{i.e.,} \quad F(x) = 1 - \exp\left[-\left(\frac{x}{\eta}\right)^\beta\right] \quad (1)$$

Note that the fact that  $\beta > 1$  implies that the component has an increasing failure rate.

The system is required to perform a sequence of missions each having length  $m$ . If the system fails during a mission, then the mission is aborted and maintenance is performed. The time required to perform maintenance is  $t_m$ , and maintenance restores the system to an "as good as new" condition. The performance of this system is measured using the average time between successful missions  $\mu$ . If the system never experiences failure, then  $\mu = m$ . However, this ideal case never occurs. Therefore, we study, using discrete-event simulation, the performance of the system under three maintenance policies.

The first system maintenance policy considered is "run-to-failure" (RTF) maintenance. Under this policy, the system is maintained only upon failure. Let  $\mu_{RTF}$  denote the average time between successful missions under this maintenance policy. Note that  $\mu_{RTF} > m$  because time is "wasted" on unsuccessful missions and system maintenance. Simulation of system performance under the RTF policy requires the manipulation of three variables: (1) the time until failure of the system ( $X$ ), (2) the number of missions successfully completed ( $N$ ), and (3) the cumulative elapsed time required to reach  $N_{\max}$  successfully completed missions ( $T_{sum}$ ). The input parameters for the simulation model are  $\beta, \eta, m, t_m$  and  $N_{\max}$ .

The second system maintenance policy considered is scheduled preventive maintenance (PM). Under this policy, an optimal, scheduled, preventive maintenance policy is applied to the system. This policy is summarized by the parameter  $\tau$ . Specifically, if a system successfully completes  $\tau$  consecutive missions, then maintenance is performed prior to the next mission. The value of  $\tau$  is determined using an embedded simulation-based optimization algorithm. Let  $\mu_{PM}$  denote the average time between successful missions under this maintenance policy. Note that, since the PM policy is optimized,  $\mu_{PM} < \mu_{RTF}$ . Simulation of system performance under this policy requires the manipulation of five variables: (1) the time until failure of the system ( $X$ ), (2) the number of missions successfully completed ( $N$ ), (3) the current number of consecutive successfully completed missions ( $N_{\text{cons}}$ ), (4) the cumulative elapsed time required to reach  $N_{\max}$  successfully completed missions ( $T_{sum}$ ), and (5) the upper limit on the number of consecutive successful missions  $\tau$ , which triggers the initiation of preventive maintenance. The input parameters for the simulation model are  $\beta, \eta, m, t_m, N_{\max}$  and  $\tau$ .

The third system maintenance policy considered is condition-based maintenance (CBM). Under this policy, scheduled preventive maintenance is replaced with a prognostic tool. The remaining life of the system is estimated at the end of each successful mission. If this estimate is less than the mission length  $m$ , then maintenance is performed prior to the next mission. We first consider "perfect prognostics", i.e. the case in which the estimate of remaining life is exactly equal to the actual remaining life  $X$ . However, a perfect prognostic is an unrealistic standard. Therefore, we also consider cases in which the prognostic test is subject to error. Under imperfect prognostics, the estimate of the remaining life is equal to  $X_{est}$  where

$$X_{est} = X + \varepsilon \quad (2)$$

and the prognostic error  $\varepsilon$  is a normal random variable having a mean of zero and a standard deviation of  $\alpha$  (note that  $\alpha = 0$  corresponds to perfect prognostics). This error creates the possibility of unnecessarily early maintenance due to underestimation of remaining life and system failure due to overestimation of remaining life. Let  $\mu_{CBM}(\alpha)$  denote the average time between successful missions under this maintenance policy. Note that  $\mu_{CBM}(0) < \mu_{PM}$ . Furthermore, note that if  $\alpha_1 < \alpha_2$ , then  $\mu_{CBM}(\alpha_1) < \mu_{CBM}(\alpha_2)$ . Simulation of system performance under this policy requires the manipulation of four variables: (1) the time until failure of the system ( $X$ ), (2) the number of missions successfully completed ( $N$ ), (3) the cumulative elapsed time required to reach  $N_{\max}$  successfully completed missions ( $T_{sum}$ ), and (4) the estimated time until failure of the system  $X_{est}$ . The input parameters for the simulation model are  $\beta, \eta, m, t_m, N_{\max}$  and  $\alpha$ .

### Experimental Design:

The next step in achieving the main objective of this research and answering the associated questions was to design an experiment for evaluating system performance over a range of choices for the system reliability and maintainability characteristics. This section details how this experiment was designed to obtain the statistics of interest.

Without loss of generality, the characteristic life of a new system  $\eta$  was set to 100. Then, four experimental factors were selected:  $\beta, m, t_m/m$  and  $\alpha$ . Ten levels of  $\beta, m$  and  $t_m/m$  were considered (Table I). All combinations of these three factors were simulated, resulting in a total of 1,000 experiments to be simulated under each maintenance policy. For the CBM policy, 16 levels of  $\alpha$  were considered for each experiment (Table I). The various levels for all these experimental factors were chosen in such a way as to envelop a wide range of operating circumstances for the system.

For each experiment 18 simulations were required: one for RTF, one for PM, and 16 for CBM. The statistics of interest collected from the simulations were the point estimates of  $\mu_{RTF}$ ,  $\mu_{PM}$  and 16 point estimates of  $\mu_{CBM}(\alpha)$  (once for each value of

Table I. Levels of the Experimental Factors

$\beta$	1.25, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00, 3.25, 3.50
$m$	10, 15, 20, 25, 30, 35, 40, 45, 50, 55
$t_m/m$	5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%
$\alpha$	0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150

$\alpha$ ). In order to ensure statistical validity of these estimates, each simulation was replicated 60 times with each replication having a run length of  $N_{\max} = 12,000$ .

### Numerical Analysis:

For all 1,000 experiments, the output of the simulation model can be used to assess the potential benefit of using CBM as opposed to PM. For each experiment, the maximum benefit resulting from the use of CBM as an alternative to PM can be estimated by

$$\frac{\hat{\mu}_{PM} - \hat{\mu}_{CBM}(0)}{\hat{\mu}_{PM}} \times 100\%$$

This value is referred to as the perfect prognostics improvement estimate and captures the percent improvement in system performance (average time between successful missions) resulting from the use of perfect prognostics ( $\alpha = 0$ ) instead of PM. Table II contains summary statistics for the perfect prognostics improvement estimate across the 1,000 experiments, and Figure 1 contains a histogram of these 1,000 estimates.

Table II. Perfect Prognostics Improvement Statistics

minimum	1.23%
median	8.71%
average	8.63%
maximum	14.55%

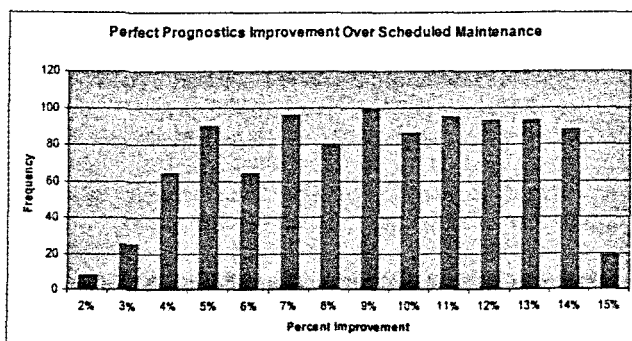


Figure 1. Histogram of Perfect Prognostics Improvement

For all 1,000 experiments, the output of the simulation model can also be used to make a more formal comparison of the three maintenance policies. These comparisons are made through the use of statistical hypothesis testing. The first set of tests attempt to prove that CBM (for each of the 15 imperfect levels of  $\alpha$ ) is superior to PM. In other words, the statistical hypothesis test is given by:

$$H_0: \mu_{CBM}(\alpha) \geq \mu_{PM}$$

$$H_1: \mu_{CBM}(\alpha) < \mu_{PM}$$

Using the output of the simulation model, these tests (15 tests for each of the 1,000 experiments) are evaluated using a two-sample *t*-test (variances not assumed to be equal) with a level of significance of 0.025. The second set of tests attempt to prove that CBM (for each of the 15 imperfect levels of  $\alpha$ ) is inferior to PM. In other words, the statistical hypothesis test is given by:

$$H_0: \mu_{CBM}(\alpha) \leq \mu_{PM}$$

$$H_1: \mu_{CBM}(\alpha) > \mu_{PM}$$

Using the output of the simulation model, these tests (15 tests for each of the 1,000 experiments) are evaluated using a two-sample *t*-test (variances not assumed to be equal) with a level of significance of 0.025. The third set of tests attempt to prove that CBM (for each of the 15 imperfect levels of  $\alpha$ ) is superior to RTF. In other words, the statistical hypothesis test is given by:

$$H_0: \mu_{CBM}(\alpha) \geq \mu_{RTF}$$

$$H_1: \mu_{CBM}(\alpha) < \mu_{RTF}$$

Using the output of the simulation model, these tests (15 tests for each of the 1,000 experiments) are evaluated using a two-sample *t*-test (variances not assumed to be equal) with a level of significance of 0.025. The fourth set of tests attempt to prove that CBM (for each of the 15 imperfect levels of  $\alpha$ ) is inferior to RTF. In other words, the statistical hypothesis test is given by:

$$H_0: \mu_{CBM}(\alpha) \leq \mu_{RTF}$$

$$H_1: \mu_{CBM}(\alpha) > \mu_{RTF}$$

Using the output of the simulation model, these tests (15 tests for each of the 1,000 experiments) are evaluated using a two-sample *t*-test (variances not assumed to be equal) with a level of significance of 0.025.

The results of these four sets of tests are summarized in Table III and Table IV. For example, when  $\alpha = 60$ , the statistical testing suggests that CBM is superior to PM for 651 of the 1,000 experiments, inferior to PM for 322 experiments, and equivalent to PM for 27 experiments. Furthermore, when  $\alpha = 100$ , the

statistical testing suggests that CBM is superior to RTF for 913 of the 1,000 experiments, inferior to RTF for 77 experiments, and equivalent to RTF for 10 experiments. Note that as  $\alpha$  increases (decreases), CBM is more often inferior (superior) to PM

Table III. Comparing CBM to PM

$\alpha$	Cases for which CBM is superior to PM	Cases for which CBM is inferior to PM	Cases for which CBM is equivalent to PM
10	971	22	7
20	910	79	11
30	860	123	17
40	814	171	15
50	737	232	31
60	651	322	27
70	569	395	36
80	502	459	39
90	437	518	45
100	388	575	37
110	340	617	43
120	301	661	38
130	259	699	42
140	233	733	34
150	208	753	39

Analysis of the results of the hypothesis testing for each value of  $\pm$  reveals apparent patterns in the individual test results. Specifically, three commonalities were observed. First, as  $t_m/m$  increases (decreases), the number of tests concluding that CBM is superior (inferior) to PM increases. Longer PM breaks have a more negative effect on system performance, therefore, it is more desirable to avoid them through the use of prognostics. Second, as  $t_m/m$  increases (decreases), the number of tests concluding that CBM is superior (inferior) to RTF decreases. This can be attributed to the fact that the only difference when comparing CBM to RTF is that under prognostics one could perform maintenance unnecessarily early due to underestimated remaining life; however, if the remaining life is overestimated, the system will just fail which is what happens with RTF. As a result, longer maintenance breaks due to underestimated remaining life should be avoided or else RTF will become more effective. Third, as  $\beta$  increases (decreases) so do the number of tests concluding that CBM is inferior (superior) to PM. This can be explained in general terms by the fact that as  $\beta$  increases, the short-term reliability of the system improves whereas the long-term reliability worsens. This characteristic leads to much unnecessarily early maintenance under CBM because the effect of prognostic error intensifies as the long-term reliability worsens.

### Concluding Remarks:

This study is based on the assumption that the failure of the system under consideration is governed by a known Weibull distribution. Therefore, this study is somewhat biased in favor of PM. Therefore, future work should consider the case in which the parameters of the Weibull distribution are subject to statistical error. In this case, the PM policy will not necessarily be optimal

and CBM will appear more effective. Furthermore, future work should consider the case in which system failure is governed by a physics-based model. In this case, CBM will be even more attractive as a maintenance policy.

This study provides a great deal of insight into how prognostic errors can impact and perhaps worsen the performance of a system. However, this study considers a system with a single-component or "black box" structure, a straightforward mission profile, a basic measure of performance, and a simple prognostic tool. Therefore, four obvious areas for further study are systems with: (1) more complex component structures, (2) more complex mission profiles, (3) more elaborate measures of performance, and (4) more realistic prognostic tools.

### Mentor comments:

Richard Cassidy, Mr. Carrasco's faculty mentor, had the following things to say about his student's work:

Although Mauricio has been involved with several research efforts during the past three years, the majority of his research activities were focused on the completion of his undergraduate honors thesis, *A Study of the Impact of Prognostic Errors on System Performance*. In this effort, Mauricio used discrete-event simulation to compare the performance of a simple system under three types of maintenance policies: (1) run-to-failure maintenance, (2) optimal, scheduled maintenance, (3) condition-based maintenance (real-time prognostics). In the case of prognostics, he considered the case of perfect prognostic information and various degrees of imperfect prognostic information.

Mauricio's work was recognized in several ways. First, he received a State Undergraduate Research Fellowship from the SILO Advisory Council. In my time at the University of Arkansas, only two of our students have obtained this fellowship. Second, he competed in two undergraduate student technical paper competitions. In the IIE Region V competition, he placed third. At the National Technical and Career Conference of the Society of Hispanic Professional Engineers, he placed first. Third, he presented a paper (in a regular session) at the 2006 Reliability and Maintainability Symposium in Newport Beach, California. For this paper, he and I received the Stan Ofthsun Award for the outstanding paper presented at RAMS, authored or co-authored by a member of the Society of Reliability Engineers. Finally, he received the 2005 Undergraduate Research Award from the Department of Industrial Engineering.



